

Copyright statement

This thesis contains material which is commercially confidential and has been registered as such with the University of Plymouth. It is not to be passed on to any third party, nor copied without the author's prior written permission until October 1st 1995.

After this date, this thesis may be supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's prior written consent.

**Intelligent fetal monitoring and
decision support in the management of labour**

by

Robert Duncan Falconer Keith

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
and Perinatal Research Group
Plymouth Postgraduate Medical School

September 1993

Abstract

Intelligent fetal monitoring and decision support in the management of labour

by

Robert Duncan Falconer Keith

The condition of the fetus during labour is inferred from the continuous plot of fetal heart rate and uterine contractions (cardiotocogram, CTG). This can be difficult to interpret which results in both unnecessary intervention and a failure to intervene when necessary causing potentially preventable neurological damage and mortality. Conventional computing approaches have not been successful in addressing these problems. This is perhaps because the correct interpretation of fetal condition requires physiological knowledge, considerable practical experience and knowledge of the specific patient.

The work described in this thesis is concerned with the investigation of artificial intelligence techniques to assist in the interpretation of fetal condition and advise on labour management. A fundamental investigation examined the performance of five types of scalp electrodes for obtaining the fetal electrocardiogram (ECG), from which heart rate is derived, and examined the factors which hamper fetal ECG data acquisition. New methods were developed to classify the important features from the CTG and included an investigation using neural networks. Other CTG features were classified using novel numerical algorithms developed closely with experts. An expert system, guided by a database of rules obtained from experts, was used to process and interpret changes in the CTG features by taking account of patient specific information. This hybrid approach was adopted to improve performance and reliability.

After two internal evaluations had found the system obtained a performance comparable with local experts, an extensive external validation was undertaken. This study involved 17 experts from 16 leading centres within the UK. Each expert and the system reviewed 50 cases twice, at least one month apart which contained those considered most difficult to interpret selected from a database of 2400 high risk labours. A novel method was developed to present all the relevant clinical information in a way which approximated the clinical situation. The reviewers scored each 15 minutes of recording according to the concern they had for the fetus and the management they considered appropriate. In this respect, this is the first reported study to examine the performance of expert obstetricians in the management of labour. A new method was derived to measure the agreement between the scores obtained and is applicable to other areas where it is required to measure the similarity between time related sequences. This study found that the experts agreed well and were consistent in their management of the cases. The system was indistinguishable from the experts, except it was more consistent, even when used by an engineer with little knowledge of labour management.

This study has shown that expertise in fetal monitoring is achievable in which case the current evidence suggests that this is not being adequately transferred to clinicians. The challenge remains to formulate a method to effectively transfer knowledge to the labour ward and thereby address the real and practical problems which face fetal monitoring today. This study demonstrates that intelligent systems could provide the vehicle to achieve this.

I dedicate this work to the memory of my father, Bradley Kenneth Keith with a hope that he always believed it possible. I know he would have had some interesting comments to make and I sadly miss the opportunity of discussing them with him.

I also dedicate this work to my mother for always being there, and to my wife Michelle for her unwavering support, patience and most of all her encouragement throughout this work.

Acknowledgements.

I would like to acknowledge and thank my supervisors,

Emmanuel Ifeakor, for his words of wisdom, constructive criticism, valued discussion, guidance and for his focused approach to research. I have learnt much over the years we've worked together and on reflection, I can appreciate the progress in my own development in which he has played a considerable part.

Keith Greene, for his valued friendship, unrestrained enthusiasm, lively discussion and motivation. If this work is destined to make a contribution to the practical world of fetal monitoring then it will be significantly attributable to his clear vision which first identified the potential and his interdisciplinary approach to research which made it a reality.

I gratefully acknowledge the significant contribution to this work by Jenny Westgate whom I consider *the* expert of experts. The success of this project is largely due to her dedication, articulation and most of all her good nature which saw her through many a tedious classification, elicitation and consultation session.

I also acknowledge the contribution of Sarah Beckley and Jon Garibaldi during the validation phase of this work. We spent many hours photocopying, cutting out and sticking together those retched CTGs; I enjoyed every minute of it. In particular, I would like to thank Sarah for her perseverance, accuracy and attention to detail in compiling the clinical details required and to Jon for his scientific mind, stimulated discussion and appreciation of wine; the Group is much enhanced by their presence. In addition, I would like to thank Nick Outram for embarking on the 'Sleeper' and taking the validation study to Scotland and to the friendly Plymouth doctors and midwives, Wame, Geoff Hughes, Jeremy Smith, Mo Harris and Anne Williams whom I managed to press gang into helping in this work.

Finally, I would like to thank the 17 experts (underlined) and their distinguished Heads of Department for their collaboration in the validation of the system. They participated with enthusiasm and gave their time freely. I enjoyed very much discussing the big issues in Obstetrics with those I visited and learnt much.

Prof Richard Beard and Dr Brian Reid, St. Mary's Hospital, London.

Prof Phil Steer and Dr Anne Deans, Chelsea and Westminster Hospital, London.

Mr Donald Gibb, King's College Hospital, London.

Prof Richard Lilford, Mr N. Johnson and Dr Imogen Montegue, St. James' University Hospital, Leeds.

Dr Kate Guthry, Hull Maternity Hospital, Hull.

Prof Martin Whittle and Dr Brian Beattie, Birmingham Maternity Hospital.

Prof David James and Dr Mitch Maharja, Queen's Medical Centre, Nottingham.

Prof P.M.S. O'Brien and Dr Mark Doyle, North Staffs Maternity Hospital Stoke-on-Trent.

Dr Paul Johnson, John Radcliff Hospital, Oxford.

Dr Karl Murphy, Queen Mother's Hospital, Glasgow.

Mr John Spencer and Dr Pauline Hurley, University College Hospital, London.

Mr S.A. Walkinshaw and Dr John Duffie, Liverpool Maternity Hospital, Liverpool.

Mr Jim Pearson and Dr Nigel Davies, University Hospital of Wales, Cardiff.

Mr Neren Patel, Mr Gary Mires and Dr C.P. Lee, Ninewells Hospital, Dundee.

Mr Tim Wheeler and Dr Pippa White, The Princess Anne Hospital, Southampton.

Mr N. Coley and Dr Racheal Sturley, Royal Devon & Exeter Hospital, Exeter.

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 study was financed with the aid of a studentship from the Science and Engineering Research Council as well as funding from, Northcott Devon Medical Foundation, Polytechnic Central Funding Council, Mason Medical Research Foundation and The Royal Academy of Engineering.

Publications; refereed.

Westgate J, Keith R D F, Curnow J S H, Ifeakor E C and Greene K R. (1990) Suitability of fetal scalp electrodes for monitoring the fetal electrocardiogram during labour. *Clin Phys & Physiol Meas* 11: 4, 297-306.

Ifeakor E C Keith R D F Westgate J and Greene K R. (1991) An expert system to assist in the management of labour. *Expert Systems World Congress Proceedings*. Orlando Florida December 16th-19th Liebowitz J. (ed). Pergamon Press, Oxford UK 4: 2615-2622.

Papers accepted for publication; refereed.

Keith R D F, Westgate J, Ifeakor E C and Greene K R. (1993) Suitability of artificial neural networks for feature extraction from the cardiotocogram during labour. *Med Biol Eng & Comp*.

Keith R D F, Westgate J, Hughes G W, Ifeakor E C and Greene K R. (1993) Preliminary evaluation of an intelligent system for the management of labour. *J Perinat Med*.

Papers in preparation.

Keith R D F, Westgate J, Beckley S, Ifeakor E C and Greene K R. The feasibility of an intelligent system to support decision making during labour. *BMJ*.

Keith R D F, Beckley S, Garibaldi J, Ifeakor E C and Greene K R. A comparison of the performance of 17 experts and an intelligent system in the management of labour in 50 cases. *Br J Obstet Gynaecol*.

Keith R D F, Ifeakor E C, Westgate J, Greene K R. Development and validation of an intelligent system to assist in the management of labour. *IEEE Trans Biomed Eng*.

Conference presentations; refereed.

Quantitative assessment of the suitability of fetal scalp electrodes for monitoring the fetal electrocardiogram during labour. 1st International Symposium on Intrapartum Surveillance, Queens Medical Centre, Nottingham. October 18-19 1990.

Suitability of neural networks for features extraction from the cardiotocogram during labour. 3rd World Symposium of Computers in Obstetrics, Gynecology and Neonatology. Anchorage, Alaska. June 6-10 1992.

An intelligent decision support tool for the management of labour. 3rd World Symposium of Computers in Obstetrics, Gynecology and Neonatology. Anchorage, Alaska. June 6-10 1992.

Poster presentations; refereed.

Development of an intelligent decision support tool for the management of labour. 26th British Congress of Obstetrics and Gynaecology. Manchester, UK. 7-10 July, 1992.

An intelligent decision support tool for the management of labour. VIIIth International Congress The fetus as a patient. Oulu, Finland. 15-18 September, 1992.

Expert systems in medicine. Society for the Study of Fetal Physiology. 20th Annual Meeting. Plymouth, UK. 16-19 May, 1993.

Signed *R Keith*

Date *10th September 1993*

"The world of reality is a world of limitations"

Gilbert Keith Chesterton (1874-1936).

Contents

Chapter 1

Introduction to the management of labour	1
1.1 The stress of labour	2
1.2 Fetal monitoring during labour	4
1.2.1 Current methods for obtaining the cardiotocograph	6
1.2.2 Interpretation of the cardiotocograph	8
1.3 The management of labour	13
1.3.1 Fetal blood sampling	14
1.3.2 Perinatal outcome	15
1.4 The impact of continuous fetal monitoring	16
1.5 Computerised interpretation of the cardiotocogram	19
1.5.1 The Maeda system	19
1.5.2 The System 8000	21
1.5.3 The Natali system	21
1.6 Summary of the current status of continuous fetal monitoring	22
1.7 A proposed hypothesis to explain the current status of fetal monitoring	23
1.7.1 The case for an intelligent system to assist in labour management	26
1.8 Objectives	27
1.9 Outline of the Thesis	27
Chapter 1 references	29

Chapter 2

Optimising the acquisition of the fetal electrocardiogram	32
2.1 Introduction	33
2.2 The physical characteristics of fetal scalp electrodes	33
2.3 The clinical performance of fetal scalp electrodes	34
2.4 Quantitative analysis of the signals obtained by fetal scalp electrodes	35
2.4.1 Identification of noise types	36
2.5 Signal analysis	38
2.5.1 Filter design	38
2.5.2 Signal saturation	40
2.4.3 Signal drop-out	40

2.5.4	Software implementation	40
2.5.5	Results analysed according to clinical grading	42
2.5.6	Results analysed according to electrode type	43
2.6	Measurement of the electrode frequency responses	44
2.6.1	Results	45
2.7	Quantitative effects of the electrode frequency responses on the electrocardiogram	46
2.7.1	Filter design using the frequency sampling method	46
2.7.2	Implementation of the frequency sampling method to obtain the electrode frequency response models	47
2.7.3	Results	49
2.8	Spectral analysis of the fetal electrocardiogram	50
2.9	Discussion and recommendations	54
Chapter 2 summary		57
Chapter 2 references		58

Chapter 3

Development of an intelligent system for labour management: Feature extraction from the cardiotocogram		60
3.1	Introduction	61
3.2	Feature extraction using artificial neural networks	62
3.2.1	The backpropagation neural network	63
3.2.2	The exclusive OR problem	70
3.3	A neural network to classify of the complete cardiotocogram	71
3.3.1	Method	71
3.3.2	Results	74
3.4	A neural network to classify the magnitude and timing of heart rate decelerations	74
3.4.1	Method	74
3.4.2	Results	75
3.5	A neural network to classify the magnitude of decelerations	76
3.5.1	Method	76
3.5.2	Optimising the number of middle layer nodes	78
3.5.3	Test of generalisation	79
3.5.4	Results	79
3.6	Discussion of the suitability of neural networks for feature extraction .	81
3.7	Baseline heart rate estimation	82

3.8	Heart rate accelerations	87
3.9	Heart rate variability	89
3.10	Contractions	92
3.11	Signal quality	96
3.11.1	Heart rate data	96
3.11.2	Contraction data	96
Chapter 3 summary	98
Chapter 3 references	99

Chapter 4

	Development of an intelligent system for labour management: The expert system.	100
4.1	Introduction	101
4.2	Knowledge elicitation	102
4.3	Knowledge representation	104
4.4	System implementation	107
4.4.1	The knowledge base	107
4.4.2	Development of the inference engine	108
4.4.3	The explanation facility	112
4.4.4	The user interface	112
4.5	System integration	115
4.5.1	Software control	116
Chapter 4 summary	118
Chapter 4 references	119

Chapter 5

Evaluation and validation of the system.	120
5.1 Introduction	121
5.2 Preliminary evaluation of the expert system	121
5.2.1 Introduction	121
5.2.2 Method	122
5.2.3 Results	122
5.2.4 Discussion	123
5.3 Preliminary evaluation of the integrated system	124
5.3.1 Introduction	124
5.3.2 Method	124
5.3.3 Results	125
5.3.4 Discussion	127
5.4 Validation of the system	128
5.4.1 Introduction	128
5.4.2 Objectives	128
5.4.3 Method	128
5.4.4 Study protocol	132
5.4.5 The system's review	135
5.5 Validation results	138
5.5.1 The distribution of scores recorded by the experts and the system	139
5.5.2 Analysis of agreement	141
5.5.3 Analysis of decisions to obtain delivery by caesarean section	149
5.5.4 Fetal blood sampling rates	161
5.5.5 Interventions in the second stage of labour	164
5.5.6 Intervention in cases with poor clinical outcome	167
5.5.7 Interventions in cases with good clinical outcome	171
5.5.8 Case discussion	172
Chapter 5 summary	177
Chapter 5 references	179

Chapter 6

Derivation of a measure of agreement and the application of statistical hypothesis tests 180

6.1	Introduction	181
6.2	Derivation of a measure to assess agreement in labour management	181
6.2.1	Derivation of a weighted agreement matrix	183
6.2.2	Measuring partial agreement	186
6.2.3	Summary of a measure to assess the agreement in labour management	191
6.3	Derivation of the statistical distribution of agreement	192
6.3.1	Derivation of the expected chance agreement	192
6.3.2	Derivation of the expected variance	198
6.4	Statistical tests to investigate agreement	199
6.4.1	The kappa coefficient measure of inter-rater agreement	199
6.4.2	The Kruskal-Wallis one-way analysis of variance by ranks	203
6.5	Application of statistical hypothesis tests	205
6.5.1	Did the experts or system obtain an average agreement in any case significantly better than expected by chance?	205
6.5.2	Did each expert and the system manage the cases similarly?	207
6.5.3	Assessment of each reviewer's overall inter- and intra-agreement	211

Chapter 6 summary	215
--------------------------------	-----

Chapter 6 references	216
-----------------------------------	-----

Chapter 7

Discussion and conclusions 217

7.1	Review	218
7.2	Discussion of the system's performance	219
7.3	Discussion of the experts' performance	223
7.4	Some personal observations	228
7.4.1	Improving the standards of fetal monitoring	229
7.5	Future development of the system	230
7.5.1	CTG feature extraction methods	230
7.5.2	Monitoring the progress of labour	231
7.5.3	On-line development	232
7.5.4	Commercial exploitation	232

7.5.5	A suggested work plan for the immediate future	234
7.6	Looking further into the future	235
7.6.1	New methods for fetal monitoring	235
7.6.2	Reasoning with uncertainty	237
7.7	Final conclusions	238
Appendix A	'C' software for analysis of the fetal electrocardiogram	240
Appendix B	'C' software implementation of a three layer backpropagation neural network	263
Appendix C	A transcript from a knowledge elicitation session	270
Appendix D	The knowledge tree	281
Appendix E	The knowledge base	295
Appendix F	An expert system written in Prolog software	304
Appendix G	An inference engine written in 'C' software	317
Appendix H	Case information and results obtained from the validation of the system	330
Appendix I	'C' software to calculate the chance expected agreement	532
Appendix J	Tables of results obtained from the Kruskal-Wallis test of Inter-agreement	536
	Tables detailing each reviewer's recommendations for caesarean section delivery	547

Chapter 1

Introduction to the management of labour.

1.1	The stress of labour	2
1.2	Fetal monitoring during labour	4
1.3	The management of labour	13
1.4	The impact of continuous fetal monitoring	16
1.5	Computerised interpretation of the cardiotocogram	19
1.6	Summary of the current status of continuous fetal monitoring	22
1.7	A proposed hypothesis to explain the current status of fetal monitoring	23
1.8	Objectives	27
1.9	Outline of the Thesis	27
	Chapter 1 references	29

1.1 The stress of labour.

Labour is a very special and emotional experience, but it is also a particularly stressful time for those involved. For the mother, changes occur in her body which are stressful and can be painful. The uterine contractions, which occur irregularly and often painlessly during the pregnancy, become progressively more frequent and more painful during labour and their timing cannot be influenced voluntarily. In the normal first stage of labour, the contractions cause the cervix to dilate from closed, to a diameter of 10 centimetres, to allow the descent of the fetus. This is likely to take several hours to complete but in hospital usually lasts no longer than 12 hours for a woman expecting her first baby. During the early part of labour the pain may not be severe but towards the end of the first stage as the cervix reaches full dilatation, it can become extremely painful indeed.

The fetus begins its descent down the birth canal during the second stage of labour as the resistance offered by the cervix has been overcome. The nature of the pain experienced here is different from that of the first stage. With each contraction, the mother takes a deep breath and bears down with all the force of her abdominal muscles. Each contraction forces the fetal presenting part down onto the pelvic floor. However, during the intervals between contractions, the pelvic floor at first pushes the fetus back up again. After being pushed down many times by the contractions and slipping back during the intervals, a time is reached when the fetus is stationary. Now, with each contraction and expulsive effort, the head moves slowly down and becomes more visible.

The stress of labour experienced by the mother is perhaps the most obvious to appreciate, but for the fetus, labour is a particularly stressful time for which it has had to evolve some sophisticated defence mechanisms to protect itself. The life-line for the fetus is the umbilical cord which carries to it oxygen-rich blood and carries away CO₂ and other by-products. During each contraction, blood flow to the placenta is inhibited and the umbilical cord can be occluded. This prevents the arrival of oxygenated blood causing short periods of hypoxaemia (lack of oxygen in the blood) and preventing the release of CO₂.

The time between contractions is therefore essential for allowing the placental circulation to be re-established. As the contractions become more frequent, the time for recovery reduces and the fetus will accumulate CO₂. If the circulation is further inhibited, then the fetus will become hypoxic (lack of oxygen in organs) which will be exacerbated with low maternal blood pressure and insufficient umbilical blood flow. The level of hypoxia is monitored by fetal chemoreceptors located in the major blood vessels which stimulate the sympathetic nervous system causing preferential blood distribution (and therefore oxygen distribution) to the essential organs; the heart, brain, adrenal glands and placenta. In addition, there is also an increase in the levels of circulating catecholamines; adrenaline and noradrenaline, which raise the heart rate and increase cardiac output. With persisting hypoxia, a critical time will

come when the oxygen level and blood flow to an organ are no longer sufficient for aerobic cellular metabolism. At this point reserves of glycogen stored principally in the myocardium and liver will be broken down as glucose by a process called glycogenolysis. This is metabolised anaerobically (without oxygen) and will provide the energy required to maintain organ function. The by-product of glycogenolysis is lactic acid which is taken up by buffering in the haemoglobin. The fetus can only maintain anaerobic metabolism for a limited time which is dependent on the level of glycogen reserves it began with and the capacity of the buffering system. This means that the growth retarded fetus, with small stores of glycogen, is particularly vulnerable to hypoxia during labour. Continued hypoxia will eventually exhaust the supply of glycogen and the production of lactic acid will overcome the buffering systems, whereupon the fetus will develop a metabolic acidosis. With a decreasing pH comes a non-linear increase in the concentration of noxious free hydrogen ions which are the agents responsible for tissue damage. Below a pH of 7.0, the enzyme systems are inhibited and further falls will lead to brain damage and eventually death (Kjellmer, 1988).

The mother and fetus are not the only parties for which labour can be stressful. It can also be a difficult time for the clinician who is charged with the responsibility of caring for not one, but two patients. In this respect, the role for the obstetrician and midwife is unlike that of any other medical speciality. The ideal conclusion for both mother and fetus is a normal vaginal delivery. This is the goal for the clinical staff too, but events can occur where an assisted delivery becomes necessary. For the mother, an operative intervention may be required if she becomes exhausted, if the presentation of the fetus is unfavourable, the fetus is large or if the cervix does not reach full dilatation. The comparative risks for the mother during labour have been dramatically reduced over the last century, which means that today she is at little risk in a hospital environment. However, for the fetus, the risks in labour are considerably greater.

The management of labour is an intensive care situation. Events can occur rapidly which require clear decision making, swift action and considerable skill. The task for the obstetrician is often to judge whether intervention is required, which for the fetus, is determined when fetal stress becomes fetal distress. On the one hand, unnecessary intervention for the fetus responding appropriately to the stress of labour is undesirable; an unnecessary caesarean section (CS) places the mother at higher risk and commits her to a major operation which reduces her chances of having normal childbirth for subsequent children. On the other hand, operative intervention is imperative for the distressed fetus who becomes at risk of sustaining neurological damage or mortality.

This dilemma has been appreciated since early times and indicates the need to monitor the condition of the fetus in some way, throughout labour.

1.2 Fetal monitoring during labour.

The goal for the obstetrician is to manage labour with minimal intervention without jeopardising the safety of the fetus. Prior to the 1960's, the only method to assess the condition of the fetus was by intermittent auscultation, where a trumpet-like device was positioned on the maternal abdomen and the beating of the fetal heart listened to. The clinician would count the number of fetal heart beats audibly detected in a period of time and express this figure as the average number of beats per minute. This would be noted in relation to contractions and persistent prolonged falls in rate acted upon.

The electrical activity of the fetal heart is described by the fetal electrocardiogram (ECG) which was first recorded by Cremer in 1906 superimposed upon the maternal electrocardiogram. During the 1960s, the emergence of relatively low cost computing power meant that the fetal ECG could be processed in real time.

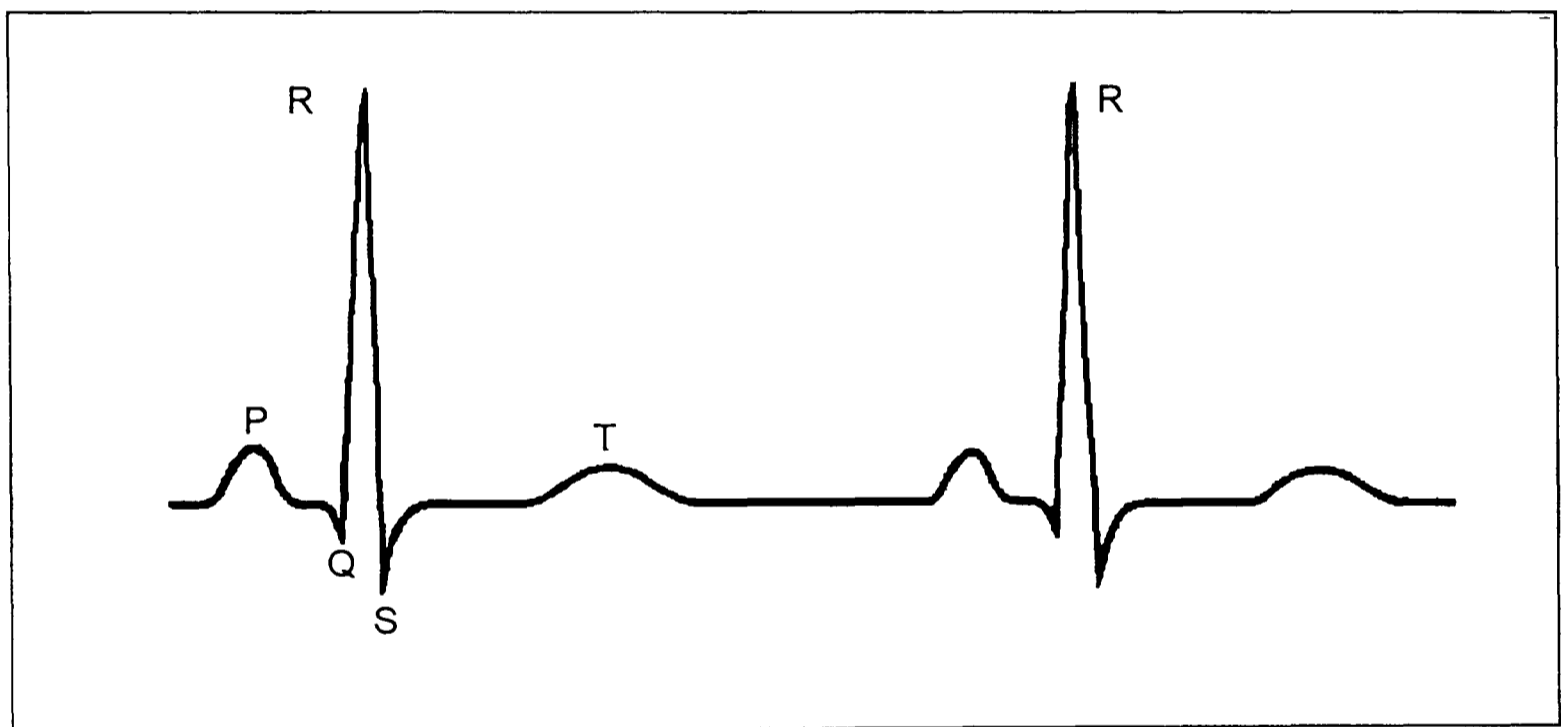


Figure 1.1: The ECG waveform.

The ECG waveform is shown in figure 1.1 and like the adult, each complex represents a heart beat and is made up of the P, Q, R, S, and T waves. The P wave represents the nervous stimulus which causes the polarised muscle cells of the heart to depolarise to form the QRS wave as they contract during the beat. The ST waveform and T wave, represent the active phase of the cardiac cycle and involve the repolarisation of the heart cells ready for the next stimulus. The fetal ECG is obtained from an electrode attached to the fetal scalp (fetal scalp electrode, FSE) and are described in more detail in chapter 2.

Each fetal heart beat can therefore be identified with each ECG complex. The R-wave is usually taken as the point of reference as it is the most prominent feature and therefore the

most accurately detected. The time between successive ECG complexes, the R-R interval, is the measurement from which the instantaneous heart rate (beats per minute) can be calculated using equation 1.1.

$$\text{heart rate} = \frac{60}{R_2(t) - R_1(t)} \quad (1.1)$$

Where, $R_1(t)$ and $R_2(t)$ represent the timings of two successive R waves expressed in seconds. The average fetal heart rate at term is approximately 140 bpm which has an equivalent R-R interval of approximately 0.43 seconds.

The development of this technique meant that it became possible to record the fetal heart rate continuously and more accurately than by intermittent auscultation which has the further disadvantage of being labour (human resources rather than childbirth!) intensive. Methods were also developed which allowed the continuous recording of the uterine contractions to be made which when plotted simultaneously with the fetal heart rate, comprise the cardiotocogram or CTG.

The equation which relates heart rate to R-R interval has interesting properties. To refer to this measure as 'heart rate' is misleading because 'rate' implies frequency. This measure is not the number of heart beats actually occurring in unit time, it simply indicates the number of heart beats which would have occurred in unit time if the single R-R interval used to derive the measure was periodic; which it is not. The timing between heart beats is seldom constant as it is continuously influenced by a highly sophisticated control system, governed by the sympathetic nervous system which seeks to reduce the interval between beats and the parasympathetic nervous system which seeks to increase the interval. The heart rate measure obtained from this equation has a non-linear, inverse relationship to the physiological variable from which it is derived. The reasons for adopting this measure rather than the actual physiological measurement are probably historical and reflect a reluctance on the part of clinicians to change from the units of measure they were used to (beats per minute), to unfamiliar units (seconds). Conceptually too, it is easy for clinical staff to visualise in terms of beats per minute. Obviously transforming the physiological variable in this way will considerably alter the profile of the resultant plot but this could have been corrected with a non-linear axis. Although this was proposed, such a scale was not adopted perhaps because a non-uniform scale is confusing or possibly it was for aesthetic reasons. The implications of introducing this transformation of the physiological variable, given the profound modification it makes to the features contained within the plot, has never been investigated.

The first clinically usable cardiotocograph was developed in 1968 by Hammacher in collaboration with Hewlett-Packard (Hammacher, 1969). Figure 1.2 shows a segment of CTG recording where the upper trace is the fetal heart rate expressed in beats per minute and the lower trace indicates the uterine contractions. The time base for the plot can be varied but is usually 1cm per minute in the UK, Europe and Australasia and 3cm per minute in the USA.

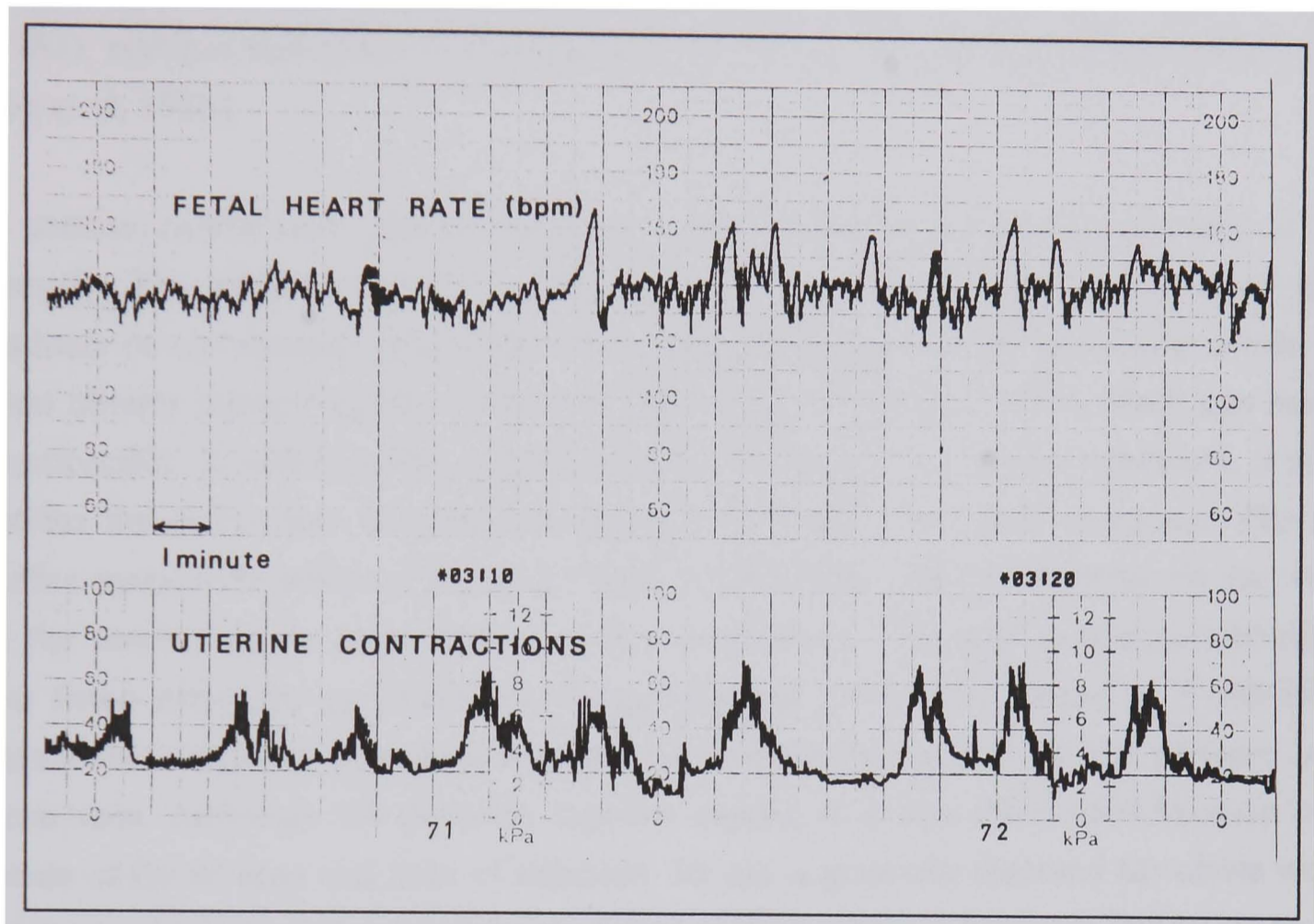


Figure 1.2: The cardiotocogram (CTG).

The CTG was rapidly introduced into clinical practice, without extensive clinical trials and quickly became the accepted standard by which the condition of the fetus during labour was monitored. Although the objectives were never clearly defined, it was hoped that continuous fetal heart rate monitoring would better identify the compromised fetus and thereby reduce or prevent stillbirth and neurological handicap resulting from birth asphyxia.

1.2.1 Current methods for obtaining the cardiotocogram.

The most accurate and preferred method for obtaining the fetal heart rate is from the fetal ECG. The ECG is obtained from an electrode inserted via the vagina to pierce the fetal presenting part following the rupture of the membranes. However, this method by definition is invasive and the signals it obtains can be difficult to analyse. This is considered more fully in Chapter 2.

The fetal heart rate may also be derived using ultrasound techniques where a transducer is secured in position on the maternal abdomen using a belt. A continuous beam of ultrasound is transmitted through the mother to the fetal heart. As the fetal heart is continuously moving, the frequency of the reflected waves undergo a Doppler shift. These changes in frequency are interpreted by the receiving equipment and the instantaneous heart rate is calculated. This method can be used prior to membrane rupture and is non-invasive. However, ultrasound is a less accurate method for obtaining the fetal heart rate (Lawson et al, 1983), and it is also easier to inadvertently record the maternal heart rate (Amato, 1983; Divon et al, 1985).

The uterine contractions can be recorded using transducers applied invasively or non-invasively. The simplest and most common method is achieved using an external strain transducer or tocodynamometer secured in place on the maternal abdomen with a belt. This system obtains inferior signals as the belt may need to be tightly fitted which can become uncomfortable. In addition, this method cannot be calibrated to measure pressure, it simply identifies the occurrence of contractions which is adequate for most purposes. The most accurate measure is achieved invasively with a fluid filled catheter inserted via the vagina into the uterine cavity upon rupture of the membranes. This uses a pressure transducer either fitted externally or located in the catheter tip. With this method, it is possible to calibrate the system to accurately measure the uterine tone as well as the pressure of the contractions. Although this provides superior signals, it is less often used because of the expense of the devices and risks of infection. Its use is generally reserved for obese women or women who have had a previous caesarean section where it becomes important to have an accurate measure of uterine pressure to minimise the risk of the contractions rupturing the uterus along the scar.

In the UK, approximately one third of women receive continuous monitoring, in the USA it is nearer half (Westgate et al, 1993). The decision to monitor continuously is taken if the patient is considered at higher risk than normal, although many would argue that all those in labour are high risk. Continuous monitoring would be commenced for example, if the woman was receiving an epidural, required drugs to augment labour, had diabetes, high blood pressure, a previous CS, was post mature, or where the fetus was large or small for dates. Those that are not continuously monitored will have intermittent auscultation.

1.2.2 Interpretation of the cardiotocogram.

The important features of the CTG are shown in figure 1.3.

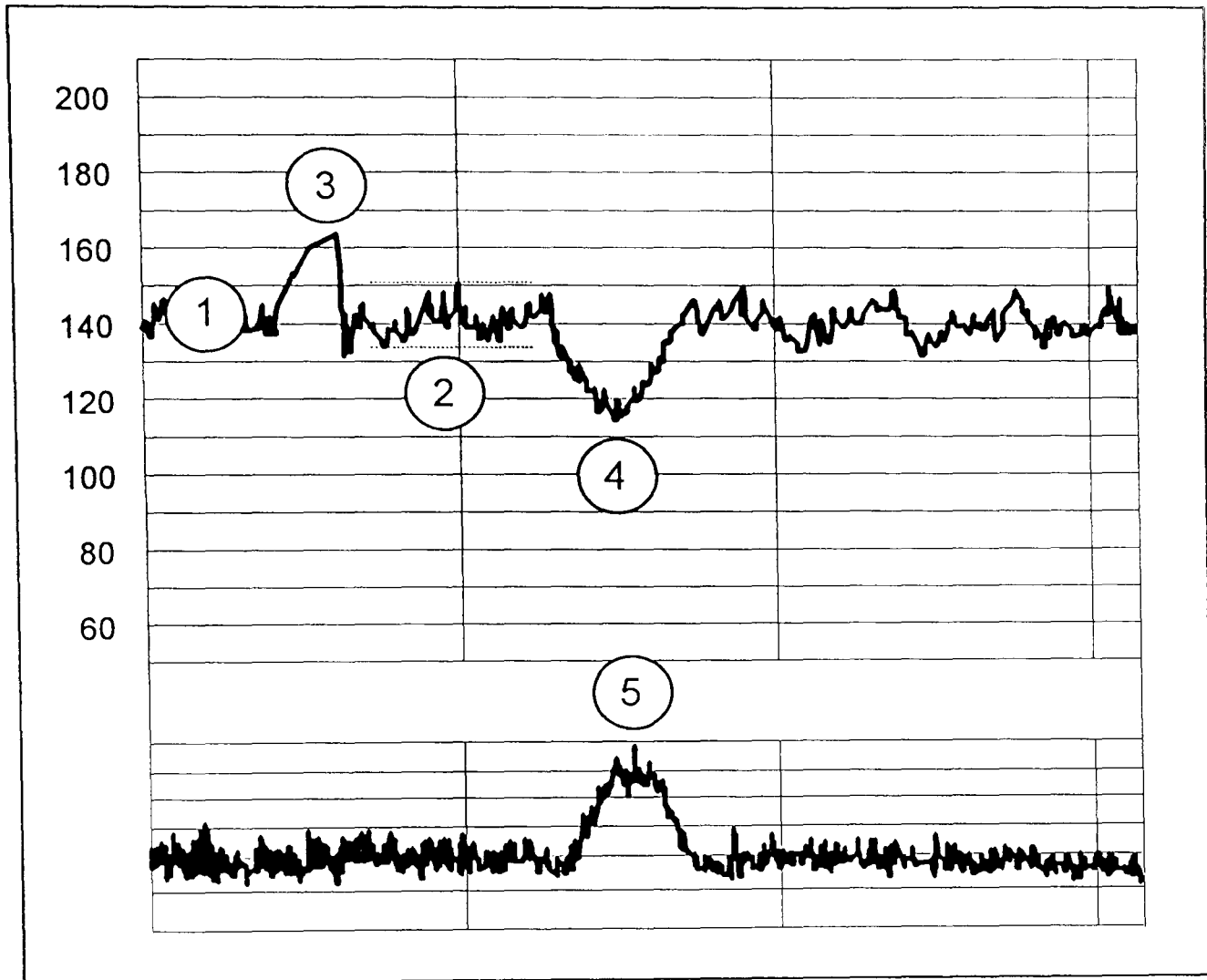


Figure 1.3: Features of the cardiotocogram.

The features labelled in figure 1.3 are,

1. Baseline heart rate; this is the heart rate value about which the heart rate pattern fluctuates. The baseline heart rate in figure 1.3 is 140 bpm.
2. Heart rate variability; this refers to the peak to peak amplitude of the high frequency perturbations about the baseline.
3. Accelerations in heart rate; relatively long term transient increases in heart rate from the baseline. The presence of accelerations is regarded as encouraging.
4. Decelerations in heart rate; relatively long term transient decreases in heart rate from the baseline which are normally associated with the contractions.
5. Location and frequency of contractions. These are important to identify because they provide information of the mothers progress and more importantly, they indicate the

times at which the fetus is subjected to the greatest stress. This gives valuable information of the fetuses capability to respond to stress and it is here that one would first expect to find signs of fetal compromise.

There have been several notable attempts by the pioneers of cardiotocography to classify the features of the CTG and standardise on its interpretation.

1. Caldeyro-Barcia's classification of decelerations.

The classification methodology suggested by Caldeyro-Barcia was first introduced in 1963 (Caldeyro-Barcia et al, 1963) and was later updated in 1966 (Caldeyro-Barcia et al, 1966). Central to this methodology was the classification of decelerations which used the relationship between the relative time of the lowest point of a deceleration and the peak of the associated contraction. Decelerations were classified into two categories:

(a) Type 1 decelerations.

This type was described as a temporary deceleration of the fetal heart rate during a contraction of the uterus, with an average lag time of 3.5 secs and standard deviation, 7.5 secs. It was said that this type of deceleration was seen after the membranes had ruptured and the fetal head was engaged.

(b) Type 2 decelerations.

This deceleration continued after the uterine contraction was complete. The average lag time was 41 seconds with standard deviation, 11 seconds. This type of deceleration was said to occur significantly more frequently when the child was entwined with the umbilical cord or where the fetomaternal blood-gas exchange was reduced.

This classification was then put into the context of the other features in the CTG to indicate the condition of the fetus. This is represented in table 1.1.

Type 1 deceleration	Type 2 deceleration	Fetal heart rate (bpm)	Fetal condition
sometimes	no	143	Normal
no	no	165	Initial fetal acidosis
no	present	> 165	Severe fetal acidosis
no	overlap	60 - 100	Critical
no	no	< 100	Critical

Table 1.1: Classification of the CTG according to Caldeyro-Barcia.

2. Hammacher's classification of heart rate variability.

Hammacher made the important contribution to CTG interpretation by attempting to classify heart rate variability, the high frequency fluctuations in heart rate about the baseline (Hammacher, 1967, 1969; Hammacher et al, 1968). The classification considered two aspects of the variability; the peak to peak amplitude and the number of oscillations per minute which are shown in figure 1.4.

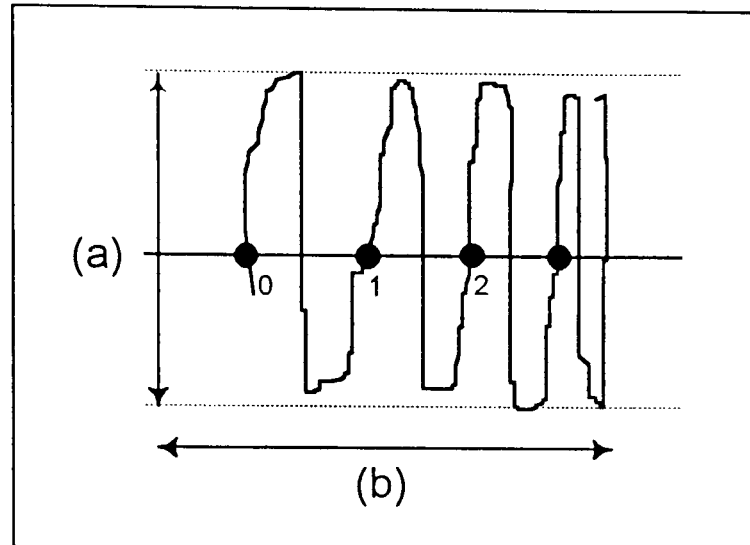


Figure 1.4: Classification of heart rate variability.

(a) Peak-peak amplitude.

The peak-peak amplitude was measured in beats per minute and was classified according to table 1.2.

Type	Amplitude (bpm)	Classification
1	< 5	Silent
2	$\geq 5 < 10$	Reduced
3	$\geq 10 < 25$	Undulatory
4	≥ 25	Saltatory

Table 1.2: Classification of heart rate variability - peak-peak amplitude.

(b) Oscillations per minute.

The number of oscillations were counted per minute and classified according to table 1.3.

Type	Oscillations per minute	Classification
a	≤ 2	Slow
b	> 2 but < 6	Middle
c	≥ 6	Fast

Table 1.3: Classification of heart rate variability - oscillations per minute.

The overall classification of heart rate variability was then obtained by considering each of the 12 possible combinations of the two measures.

Although Hammacher proposed that the classification of variability be comprised of 2 indicators, today it is generally the amplitude of variations that is taken as important. The number of oscillations are seldom used because they are difficult to assess visually and are thought not to contain additional information.

3. Hon's classification.

Ed Hon perhaps made the greatest single contribution to fetal monitoring with his pioneering work in the 1950's and 1960s. He first devised a method for classifying the CTG in 1958, which has been refined and updated throughout the years (Hon, 1958, 1959, 1963, 1967, 1968; Hon and Quilligan, 1967).

(a) Baseline heart rate.

This was classified into 5 categories as described in table 1.4,

Baseline heart rate (bpm)	Classification
> 180	Marked tachycardia
$> 160 - 180$	Moderate tachycardia
$120 - 160$	Normal
$100 - 119$	Moderate bradycardia
< 100	Marked bradycardia

Table 1.4: Hon's classification of baseline heart rate.

These ranges have subsequently been modified so that today, the current wisdom is that a marked bradycardia is < 90 bpm, moderate bradycardia is of the range, > 90 to < 110 bpm and the normal range is from 110 - 160 bpm.

(b) Decelerations.

Decelerations were classified into 3 main categories; early, late and variable, according to their shape and timing in relation to contractions.

Early decelerations; when the deceleration was congruent to an accompanying uterine contraction and the onset of deceleration and contraction were simultaneous, the deceleration was classified as an early deceleration.

Late decelerations; when the deceleration was congruent with an accompanying uterine contraction, but the onset of the deceleration lagged the onset of the contraction, the deceleration was classified as a late deceleration.

Variable deceleration; when the shape of the deceleration was different from the shape of the accompanying contraction, the deceleration was classified as a variable deceleration. In addition, variable decelerations could be further classified as severe if they dropped below the baseline by greater than 60 bpm, dropped below 60 bpm, or lasted longer than 60 seconds (the so called 'rule of sixties').

General comments.

These various classification methods have not standardised CTG interpretation. In fact it could be argued that the opposite was true. For example, some obstetricians will refer to decelerations as early, late and variable, others will refer to them as type 1 or type 2 and some will mix them up. However, the classification of decelerations proposed by Caldeyro-Barcia were recently described as "archaic", whereas the classification according to Hon were considered more suitable (Neilson, 1993). Several attempts have been made to improve CTG classification (Caldeyro-Barcia et al, 1974; Zuspan et al, 1979; Rooth et al, 1987) and guidelines have been recommended (American College of Obstetricians & Gynecologists, 1972), but again without the desired success. However, it has been found that these classifications and guidelines have been difficult to apply which is perhaps because they are subjective. For example, how late does a deceleration have to be to be classified as late? The CTG is a dynamic waveform and seldom maintains a particular classification. Abnormalities are frequently not persistent. These methods also do not assist the clinician with their subsequent management, for example, during a bradycardia (baseline < 90 bpm) in the first stage of labour; should the clinician always deliver immediately by emergency CS?

The classification of Hon and the classification of variability by Hammacher (amplitude only) generally form the basis of the methods applied today. They provide a useful description of the features but they do not allow a straight forward interpretation of the CTG. The heart rate changes seen in the CTG are often physiological in origin rather than pathological and as such, mirror the complex fetal compensatory mechanisms at work. A variable deceleration, for example, can result from the compression of the umbilical cord during a contraction which reduces the venous blood flow. Consequently, there is a decreased demand for blood to be pumped, so the heart rate reduces. When the contraction is over and the blood flow returns, the heart rate rapidly returns back to normal. Problems are not necessarily indicated by the presence of decelerations, but can develop as a consequence of them; if the decelerations become too frequent then the time between contractions becomes insufficient to re-establish blood flow which begins the familiar cycle which ultimately leads to fetal distress. In addition, consider heart rate variability which provides important information regarding the functioning of the central nervous system. These fluctuations are caused by the mechanisms which seek to increase heart rate and the mechanisms which seek to decrease it, antagonising each other. However, a reduced variability may not be abnormal, because surprisingly perhaps, the fetus has sleep cycles during labour in which the heart rate variability is naturally reduced.

The bottom line is that for the intrapartum CTG, an abnormal feature is not an abnormal occurrence. The correct interpretation of the CTG depends on an appreciation of the fetal physiological factors that influence heart rate changes which are inter-related and require considerable skill and experience to distinguish from pathological changes. This point is absolutely fundamental to understanding the difficulties associated with fetal monitoring.

1.3 The management of labour.

It has been mentioned that labour is an intensive care situation in which events can occur very rapidly. It is therefore essential that the clinical staff charged with the care of the patient are extremely vigilant.

The patient is placed in the continuous care of the midwives who will call the junior doctor on duty if there are perceived problems. It is then the junior doctor who decides whether the perceived situation warrants a consultation with a more senior colleague which is usually the Registrar, or Senior Registrar. In turn the Registrar can seek further advice from the Consultant. This management structure is represented in figure 1.5.

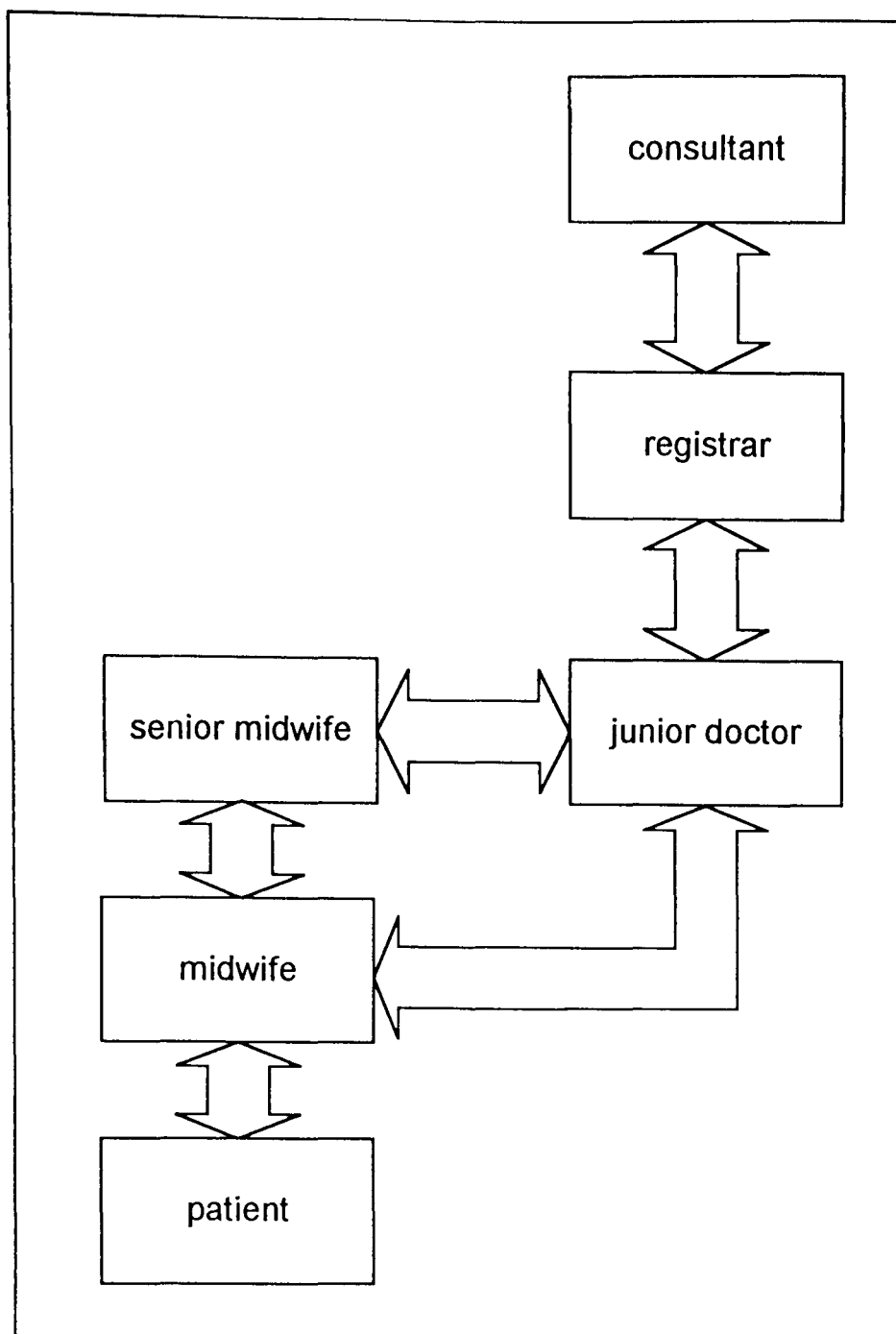


Figure 1.5: The structure of clinical staff for labour management.

It can be appreciated that this structure has potential weaknesses. In the first place, the patient is furthest removed from the most experienced clinician, the Consultant. But perhaps the most serious weakness of this structure, is that the care of the patient depends on the most junior of staff, the midwife and junior doctor who may have little experience. They are relied upon to first form an assessment of the patient (mother and fetus) on which they decide whether more senior staff should be called.

1.3.1 Fetal blood sampling.

Where the evidence from the CTG is inconclusive but where it gives rise to significant concerns for the fetus then additional information can be obtained from a fetal blood sample (FBS). The pH of the blood gives an indication of the metabolic state of the fetus. The blood is obtained by passing an endoscope via the cervix and then using a very small blade to make an incision into the fetal scalp, or bottom if in the breech presentation. A drop of blood is drawn into a long fine capillary tube which can be analysed using a specialist piece of equipment. If the pH is greater than 7.25 then there is no evidence to suggest

hypoxia at this time. A pH of 7.20 and lower is significant in the first stage of labour. In the second stage of labour when there are concerns for the fetus, a FBS may not be obtained if the baby can be delivered simply with forceps, but it may still be used when the fetal presentation is unfavourable.

Whilst continuous monitoring using the CTG is widely used in the UK, the facilities required for fetal blood sampling are not always available (Johnson et al, 1990).

1.3.2 Perinatal outcome.

It is important to assess the condition of the baby at birth (referred to as perinatal outcome) to establish how well it has coped with labour. This is particularly difficult to do as no single physiological variable provides an accurate indication. Physiologists and obstetricians do seem to be standardising on three variables as being important in defining perinatal outcome. Two of these variables are the cord arterial pH and Base Deficit in the extra-cellular fluid (BDecf), which indicate whether the baby is acidotic (pH) and whether the acidosis has a significant metabolic component (BDecf). These are both measured from a sample of blood obtained from the umbilical artery after birth but before the detachment of the placenta. The third variable is the Apgar score which is a subjective indication of the condition of the baby measured at 1 minute, 5 minutes and sometimes 10 minutes after birth. This is a score in the range of 0 to 10 where 10 indicates the extremely vigorous baby and 0 indicates the baby without visible signs of life. The scoring method is shown in table 1.5.

Vital Sign	Score		
	0	1	2
Heart rate	absent	less than 100 bpm	more than 100 bpm
Respiratory effort	absent	slow, irregular	good, crying
Muscle tone	limp	some limb flexion	active
Response to stimulus	nil	grimace	vigorous cry
Colour	blue, pale	body pink, limbs blue	pink

Table 1.5: The Apgar scoring system to assess perinatal outcome.

All hospitals in the UK record Apgar scores but not all have the facilities to record blood gas analysis and fewer still have the inclination (Johnson et al, 1990), but this situation is changing as it is realised that Apgar scores, by themselves, are not sufficiently accurate (Westgate, 1993). Indeed, none of the three variables are accurate by themselves but when used together they can obtain a reasonable indication of fetal condition, especially at the extremes of the measures. For example, most would accept that a baby born vigorous (Apgar high) with an arterial pH > 7.20 (without acidosis) has no evidence of distress and that a baby born depressed (low Apgar) with a significant metabolic acidosis (pH < 7.05, BDecf ≥ 12) could be described as compromised. As one moves from the extremes and into the grey area, then one becomes less certain of the assessment. Consider babies born with acidosis for example (pH < 7.05). All babies with pH < 7.05 but without a significant metabolic component may not be compromised. Indeed in some cases, the very fact that a baby has mounted such a defence to cope with the stress of labour could indicate the contrary. On the other hand some babies born with pH > 7.05 can still be compromised. These may have become compromised as a result of a severe insult sometime before labour began or may be simply deficient in glycogen (growth retarded babies) and unable to sustain anaerobic metabolism (which produces lactic acid and reduces pH).

1.4 The impact of continuous fetal monitoring.

Continuous electronic fetal heart rate monitoring (EFM) was introduced into clinical practice after research had shown, outside the rigours of a randomised trial, that heart rate changes correlated with indicators of fetal hypoxia; specifically, intrapartum fetal death, fetal blood pH and Apgar scores (Kelly et al, 1973; Paul and Hon, 1974). The first prospective randomised trial comparing continuous intrapartum fetal monitoring with intermittent auscultation in the clinical situation, was undertaken by Haverkamp in 1976, sometime after its wide spread introduction (Haverkamp et al, 1976). This study was not able to find a significant benefit in the clinical setting when using EFM compared with auscultation carried out at 15 minute intervals during the first stage of labour and at 5 minute intervals during second. In addition, a subsequent study of the children involved in the trial was unable to show any long term benefits of EFM either (Langendoer et al, 1980).

Critics have pointed out that the numbers involved in this study were small (242 in EFM group, 241 in intermittent auscultation group) and predicted that with larger numbers, the benefits of EFM would become apparent (Parer, 1979). However, since this first study, other notable studies from around the world have compared continuous monitoring with intermittent auscultation and have broadly come to the same conclusions. The combined total of four of the most significant studies was (890 + 13,000 + 34,995 + 246) 49,131 patients (Wood et al, 1981; MacDonald et al, 1985; Leveno et al, 1986; Luthy et al, 1987).

Haverkamp and colleagues also found that continuous monitoring increased operative intervention. The inclusion of fetal blood sampling reduced, but did not remove the excess intervention in the continuous fetal monitoring group (Haverkamp et al, 1979). These studies have been substantiated. It has been found that when used on its own, EFM quadruples intervention (Grant, 1992a) and when used with fetal blood sampling it doubles unnecessary intervention compared with intermittent auscultation (Grant, 1992b). In the USA, the national CS rate was 4.5% of all deliveries prior to the introduction of continuous monitoring, after which this figure rose to 12.5% in 1975 and in 1979 was close to 20%. Currently in the USA, the figure is over 30% in most hospitals. In the UK, the current average is between 10% - 15% but can be as high as 25% in some hospitals. There is also a knock on effect for those women who have a CS because they become far more likely to have another CS for subsequent children.

The impact of continuous fetal monitoring has also increased litigation. A study examined the cases contained in the files of Action for Victims of Medical Accidents which had resulted in stillbirth, perinatal or neonatal death, or long term mental or physical handicap (Vincent et al, 1991). Out of 34 cases, the CTGs in over half were misinterpreted or not acted upon. In 17 cases, junior doctors failed to recognise fetal distress and managed a delivery they did not have the experience to deal with. In the USA, litigation has reached disproportionate levels and the fear of litigation contributes to the high levels of unnecessary intervention.

The litigation climate which surrounds fetal monitoring during labour was best described by an eminent defence lawyer during a recent closed meeting convened at the behest of the Department of Health. The lawyer's argument is perhaps a little unscientific in places but was based on his considerable experience and knowledge of the area. It took the following form:

- The conventional wisdom is that 1500 babies are born each year in the UK with cerebral palsy (The Spastics Society).
- The total population of cases with cerebral palsy is in excess of 50,000, although 70,000 is frequently quoted.
- It is conventional wisdom amongst neurologists that 8% - 15% of these patients have a lesion attributable to birth injury but that such figures tend to be reached on the basis of surveys of clinical notes, which can be self-serving. In the cold light of day before a hostile forum it can be much harder to prove that there is no peri-partum event to explain why a particular child has been born damaged. It was considered that the overwhelming majority of cases have a claim either against the obstetrician or against those who subsequently looked after them until proven otherwise. This did not mean that the case

was not resistible, but that it was good enough to get a Legal Aid Certificate.

- It was considered that on balance, half the cases with cerebral palsy were likely to make a claim and that of this half, half would be successful.
- A baby born today will not expect to get to court for 5 years. In which time, with modest inflation, the average level of compensation will be £1 million.
- Therefore, the total settlements for the total population of cases with cerebral palsy is,

$$\frac{70,000}{4} \times \text{£1million} = \text{£17.5billion}$$

- and the settlements for the new arrivals per year is,

$$\frac{1500}{4} \times \text{£1million} = \text{£350million}$$

- The lawyer concluded that at this current rate, within a period 5 to 10 years, the NHS would be spending more on the damages and costs of lawyers in dealing with the consequences of adverse obstetric results, than it was spending on the whole of obstetrics.
- As an equally reasonable second stage proposition, the lawyer suggested that within about 25 years, as much would be spent on these settlements as is currently spent on the whole of the NHS.

The important factor here is that this money for compensation comes directly from the NHS budget and not as one may suspect, from contingency funds.

Another worrying aspect of fetal monitoring is that experts do not agree well (low inter-agreement) and are inconsistent (low intra-agreement) in their interpretation of the CTG. This has been the finding of studies which examined the antenatal CTG where the fetus is unstressed (Gagnon et al, 1993; Lotgering et al, 1982; Trimbos and Keirse, 1978). In a study supported by the European Community (Donker, 1991), 13 cases of which 9 were intrapartum recordings, were considered by 21 European experts assembled on a Greek island. This study specifically examined the experts' ability to describe the CTG features present in selected segments from the cases. At the end of the segment, the experts were asked to predict the perinatal outcome. The study found little consistency or agreement in the experts' classification of most of the CTG features, nor in their prediction of outcome. Another study involved the retrospective audit of selected cases by the Consultants of a leading teaching hospital in the UK (Barratt et al, 1990). This study too found that there

was significant disagreement between the auditors in the recommended management of these cases. In addition, when presented with identical information at a different time, the auditors were inconsistent in 25% of the cases. Whilst the design and statistical analysis of some of these studies (Gagnon et al, 1993; Donker, 1991) could be criticised, their general findings would be difficult to refute; the CTG is an imprecise measure of fetal compromise.

1.5 Computerised interpretation of the cardiotocogram.

It was because of the difficulties associated with the interpretation of the CTG that significant attempts were made to computerise the process. With such an approach, it was considered that at least an objective and consistent interpretation would be obtained.

1.5.1 The Maeda System.

Kazuo Maeda of Japan began forming his ideas for automatic assessment of the fetus during labour in 1969 (Maeda et al, 1969). A computerized system was developed which extracted features from the CTG and scored them according to their perceived relevance in diagnosing fetal distress. If the combined score passed a pre-defined threshold then fetal distress was indicated.

The system first established the baseline heart rate and heart rate variability. Accelerations were defined as transient increases in heart rate with an amplitude of at least 12 bpm for longer than 12 seconds. Decelerations were defined as transient decreases in heart rate of at least 20 bpm for 30 seconds. Further parameters were then calculated to enable a classification of the type of deceleration to be made. The presence of features were then scored according to table 1.6 (Maeda, 1990).

Feature	Measure	Score
Baseline heart rate	110-160	1
	160-180	1
	< 110	3
	> 180	3
Deceleration duration	> 60 s	3
Deceleration amplitude	> 50 bpm	2
Deceleration minimum heart rate	< 100 bpm	2
Recovery time of deceleration	> 40 s	3
Lag time of deceleration with contraction	> 40 s	3
Shape of deceleration	Deceleration with no acceleration	2
	W-shaped Deceleration	4

Table 1.6: Maeda's classification of the cardiotocogram.

The sum of the fetal heart rate scores was found for each 5 minutes of monitoring. This score was then used to calculate the fetal distress index using the classification described in table 1.7.

Parameter	Measure	Index
Fetal heart rate score	10 - 19	1
Fetal heart rate score	> 20	2
Baseline heart rate	> 160 for > 30 minutes	1
	110 - 120 for 10-30 minutes	1
	110 - 120 for > 30 minutes	2
	< 100	3
Variability loss		1
Decelerations	Late	1
	Deep and persistent	3

Table 1.7: Maeda's interpretation of the CTG.

This index was assessed for every 15 minutes of monitoring. A score of 0 was considered normal and suggested no fetal compromise. A score of 1 indicated that further monitoring was required, 2 points were considered 'suspicious' and 3 points or more indicated a diagnosis of fetal distress (Maeda, 1980). Variations on this scoring system have also been published in recent times (Bedrich and Zdenek, 1990).

1.5.2 The System 8000.

The System 8000 (Dawes et al, 1991a) is a monitor designed for the antenatal assessment of the fetus and has been developed for commercial use. The interpretation of the antenatal CTG is recognised as a far easier task as the fetus is not subjected to the stress of labour. The same features are identified, but generally speaking, the presence of any abnormality gives cause for concern.

The development of this system has been well reported with specific mention of the methods used to extract the CTG features. The algorithms for baseline heart rate estimation, from which other features are defined, have been detailed (Dawes et al, 1981) and compared (Dawes et al, 1982). Methods to classify variability and their relevance to the antenatal trace have also been examined (Henson et al, 1983; Dawes et al, 1990, 1991b). Accelerations were defined as transient increases above the baseline by more than 10 bpm for at least 15 seconds. Decelerations were defined as a fall in heart rate of more than 20 bpm for at least 30 seconds, or a fall of 10 bpm for greater than 1 minute (Dawes et al, 1991a). There are no scoring systems associated with this monitor. Its purpose is to objectively extract and quantify features seen on the antenatal trace and display the findings.

1.5.3 The Natali system.

The development of the Natali system began in the 1970's by Krause, in what was East Germany (Krause, 1990). The approach here was similar to Maeda's except that an internal contraction transducer was used rather than an external belt. Detailed measurements of parameters extracted from the cardiotocogram were made and combined to calculate an index obtained from a discriminative function based on a mathematical model.

Each heart rate parameter had an associated coefficient ($a(1)$ to $a(10)$) associated with it which was proportional to the parameters' supposed importance in assessing fetal condition. A calculated index in the range 0 - 60 corresponded to a normal CTG, between 61 - 100 indicated declining fetal condition and over 100 that there were signs of hypoxia whereupon an alarm sounded.

These systems consider the CTG in isolation and do not attempt to support decision making. Consequently, they represent sophisticated alarm systems which draw the attention of clinical staff to the abnormal CTG. Although these systems have been developed over a period of 10 - 15 years they have not been submitted to a clinical assessment from which their effectiveness could be established and a system to interpret the CTG during labour is not commercially available.

1.6 Summary of the current status of continuous fetal monitoring.

1. The CTG has not lead to an improvement in the steadily falling perinatal mortality rate and there has been no change in the incidence of cerebral palsy and handicap over the same period (Jarvis et al, 1985).
2. The range of false positive abnormalities present in the CTG are thought to range from 33% (Gabert et al, 1976) to 80% (Schifrin and Dame, 1972) and make the CTG difficult to interpret.
3. There are significant levels of unnecessary operative intervention (Haverkamp et al, 1979).
4. There are a significant number of cases with potentially preventable neurological damage and mortality associated with birth asphyxia (Vincent et al, 1991).
5. The litigation and compensation in cases which are judged to have been damaged or lost as a result of birth asphyxia is reaching alarming proportions (Vincent et al, 1991).
6. In 10 to 15 years of development, a computer solution has not been demonstrated to agree with experts or accurately identify fetal compromise during labour.
7. Experts find it difficult to agree and be consistent in their interpretation of the CTG alone (Donker, 1991).

The evidence appears irresistible and it is not surprising that some have formed the opinion that there is little value in continuous fetal heart rate monitoring. And yet, the CTG has not been abandoned. On the contrary, it is used routinely in virtually all maternity units in the western world.

1.7 A proposed hypothesis to explain the current status of fetal monitoring.

The first important point to bear in mind when developing a possible hypothesis to explain the current status of fetal monitoring, is that the machines used to obtain the CTG are recorders and not monitors. The monitor is the clinician who interprets the recording. This is an important distinction because it implies that the standards in fetal monitoring are limited not only by the variables that are used to indicate fetal distress but also by the ability of the clinical staff who interpret them.

The assumption implied throughout the development of continuous fetal monitoring is that the interpretation of the CTG is easy. This is so because it has been left to the most inexperienced staff to interpret the CTG for which there is currently no formal training. How would the current status of fetal monitoring be interpreted if it were instead assumed that the interpretation of the CTG was difficult and could only be mastered by a few of the most experienced clinicians? Here, if CTG interpretation were left to the most inexperienced, then would it not be likely that incorrect interpretations would be made which would lead to both unnecessary interventions and a significant incidence of birth asphyxia. In addition, this assumption may also explain why experts involved in the pioneering work of the CTG were able to establish a correlation with hypoxia, and why this has been difficult to reproduce on the labour ward. One can only speculate on the results of a trial which compared the management of labour solely by those most experienced with current practise.

That is not to say that the CTG contains all the information required to identify the compromised fetus because the evidence from the studies which examine expert inter- and intra-variability suggest that it does not. It could be that the reason the experts were not able to agree highly and be consistent was that they were not provided with all the information they would normally have and that the studies were not conducted in a similar way to the clinical situation. A more cynical view would be that the reason the experts did not agree and were not consistent was because they were not experts. This becomes more plausible if the interpretation of the CTG is assumed to be very difficult; here, it is likely to be those with most experience and still practising who are the true experts rather than those at the pinnacle of their careers who may be far removed from the labour ward. However, the development of the alternative hypothesis does not depend on this cynical view, but it does depend on the two following assumptions,

1. The interpretation of the CTG is difficult and requires considerable experience and physiological knowledge to extract all the relevant information.

2. The CTG is an essential, but not the only source of important information which determines the appropriate management of labour.

With these assumptions a model to illustrate the process of labour management can be derived which could explain the current status of fetal monitoring without concluding that it has little value.

1. The goal for fetal monitoring is to accurately distinguish the normal group of fetuses (those responding appropriately to the stress of labour) from the abnormal group (those that become compromised). This goal is illustrated in figure 1.6 which shows the 2 groups separated. If these groups overlapped to form a 'grey area', then it would indicate that for some cases, the condition of the fetus was uncertain.

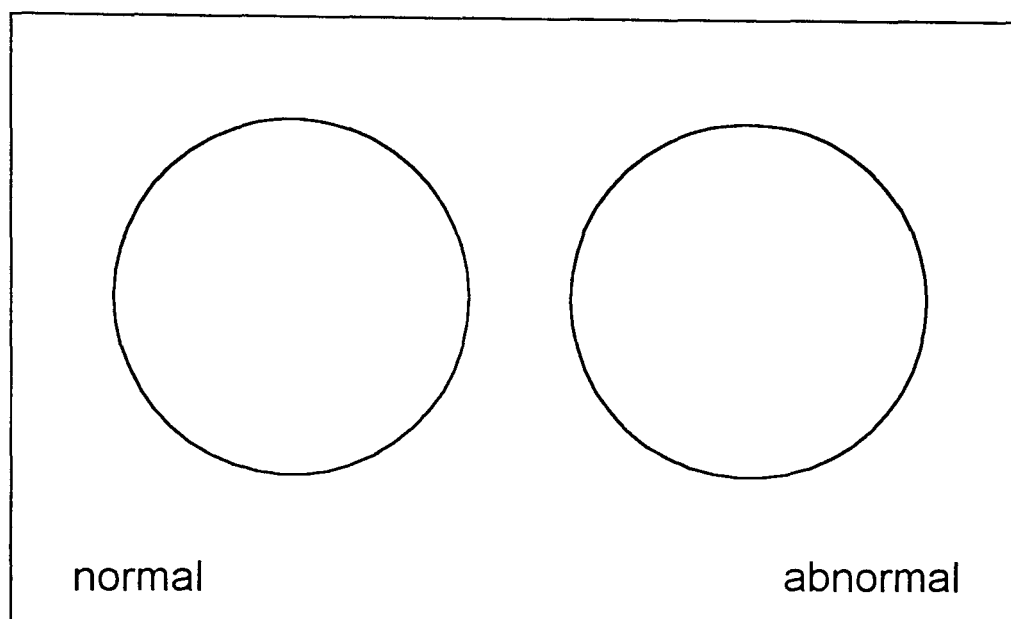
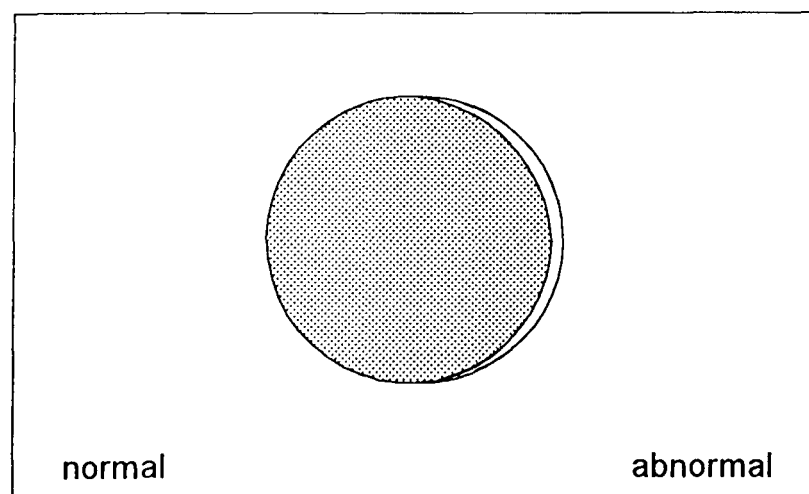
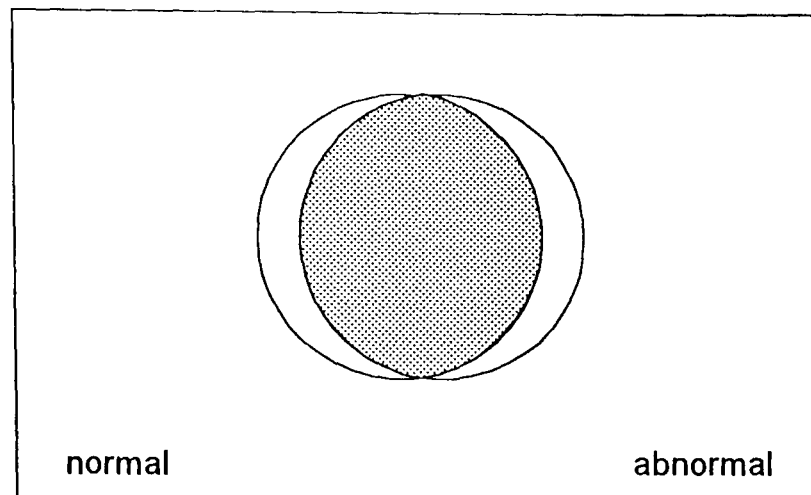


Figure 1.6: The goal for fetal monitoring.

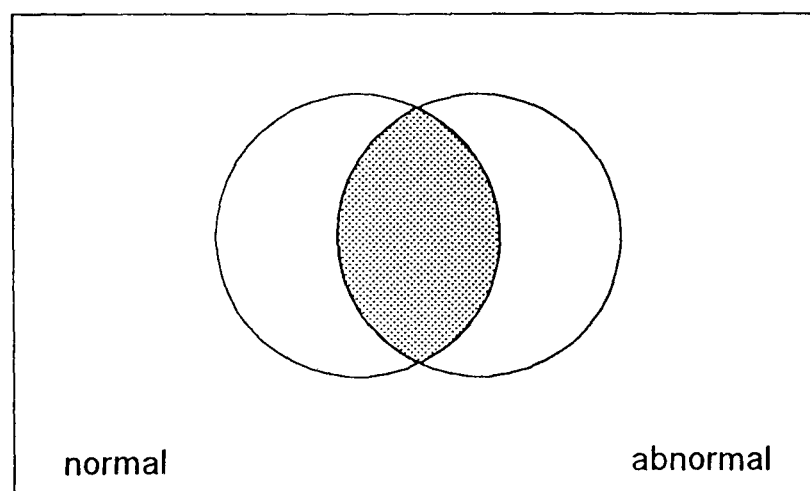
2. Without any form of monitoring only the most abnormal of cases could be expected to be identified. Therefore without monitoring, the vast majority of cases which include all the normals and virtually all the abnormal, would remain in an overlapping grey area where the condition of the fetus could not be determined accurately.



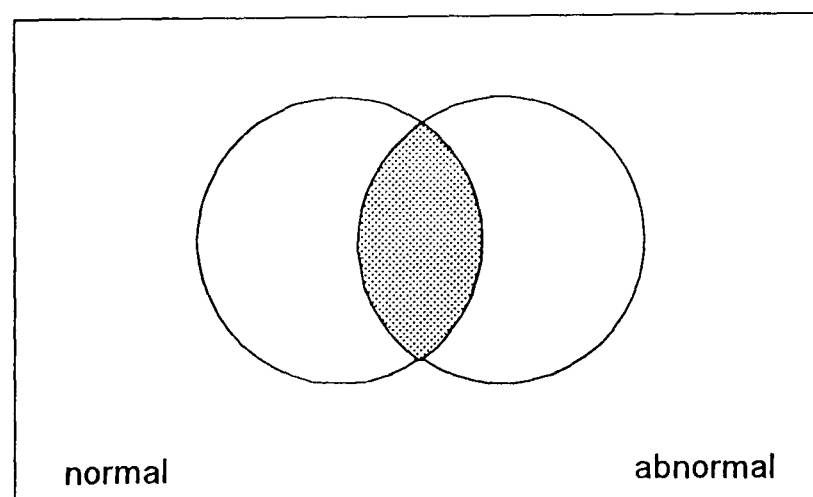
3. If the CTG were used by the untrained and inexperienced then they would be expected to identify some of the normals and some of the abnormal, but for the vast majority of cases, their interpretation would be inaccurate.



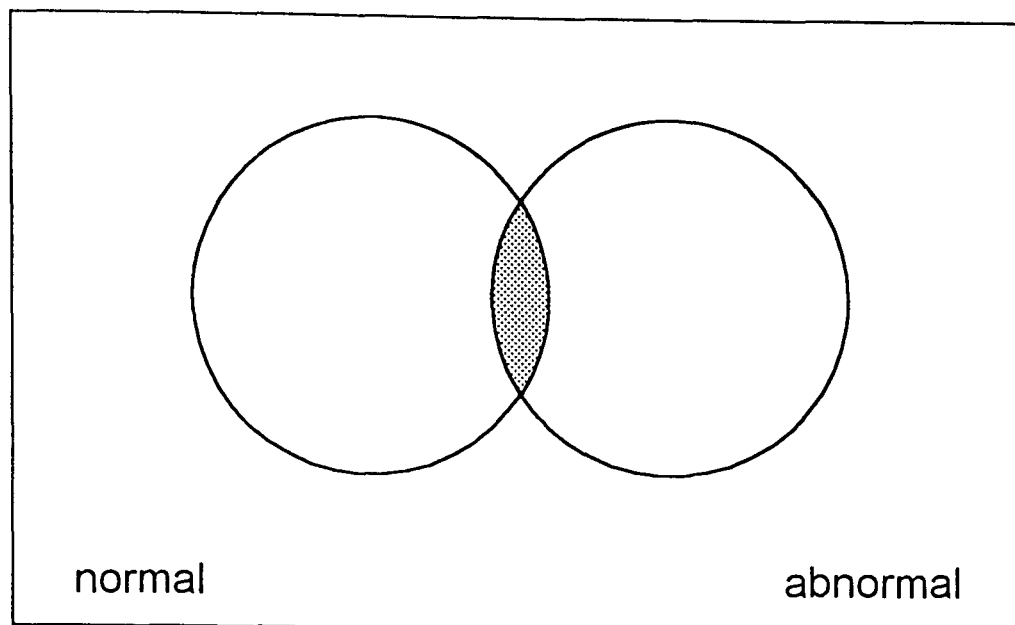
4. Experts, using the CTG alone, would be able to better separate the groups using their experience and physiological knowledge. However, there would still remain a significant 'grey area' where the condition of the fetus could not be accurately determined.



5. If experts had the complete patient information which included the patients obstetric history and details of the specific labour, then the groups would be further separated.



6. Finally, If experts interpreted the CTG using their experience and physiological knowledge, had the complete patient information and could obtain additional information with fetal blood sampling then the groups could become almost separated leaving only a small proportion of cases where the condition of the fetus is uncertain.



1.7.1 The case for an intelligent system to assist in labour management.

This model offers an alternative interpretation of the current status of fetal monitoring which does not denigrate its value.

This model predicts that a conventional computing approach which attempts to identify fetal compromise using only information obtained from the CTG will be unsuccessful. This is because even if it were able to match the expert in this, it would not be able to accurately identify the compromised fetus because it would not consider all the important information.

Importantly, this model suggests that the compromised fetus can be accurately identified by an expert in CTG interpretation, with knowledge of fetal physiology, the specific patient and an option to obtain additional information from a fetal blood sample. If this is the case then experts, when given this information, should tend to agree and be consistent in the management of labour. In addition, if this knowledge could be formalised then the model also suggests that an intelligent system which embodied this expertise could agree with experts and could help clinical staff in their management of labour.

1.8 Objectives.

The objectives for this study were to establish whether,

1. The expert knowledge required for labour management could be accurately formalised.
2. Experts could agree and be consistent in the management of labour.
3. An intelligent system could be developed to obtain a performance in labour management comparable with experts.

1.9 Outline of the Thesis.

Chapter 2 examines the acquisition of the fetal electrocardiogram from which heart rate is derived. It has been found that this signal is difficult to obtain reliably and is often of poor quality which can hamper monitoring. The purpose of this work was to identify the factors which contribute to this and suggest methods for improving the quality of the signal.

Chapters 3 and 4 describe the development of an intelligent system for the management of labour. Chapter 3 considers the suitability of numerical algorithms and artificial neural networks for feature extraction from the cardiocogram. Chapter 4 describes the development of the expert system which after evaluation, was integrated with the feature extraction methods to form the complete intelligent system.

Chapter 5 describes the evaluation and validation of the system. The system was evaluated at two key stages in its development by internal experts before being subjected to a validation study which involved 17 experts from 16 of the leading centres in fetal monitoring in the UK. This study examines whether experts can agree and be consistent in the management of labour and whether the system can attain a comparable performance.

Chapter 6 presents the detailed mathematical and statistical techniques used to analyse the results of the validation study described in chapter 5. In particular, it describes the derivation of a method to calculate the agreement in the management of labour between two reviewers and further shows how this may be applied to measure the statistical significance of the results.

Chapter 7 presents a review of the thesis and discusses the validation study results. It also suggests the likely role for the system in a strategy to improve fetal monitoring and considers future development.

Chapter 1 references.

- Amato J C. (1983) Fetal heart rate monitoring. *Am J Obstet Gynecol* **147**: 967
- American College of Obstetricians and Gynecologists. (1972) Technical Bulletin No. 32: Guidelines for monitoring, terminology and instrumentation. Harper and Row Hagerstown PA USA.
- Barrett J F R, Jarvis G J, MacDonald H N, Buchan P C, Tyrell S N and Lilford R J. (1990) Inconsistencies in clinical decisions in obstetrics. *Lancet* **336**: 549-551.
- Bedrich S R P and Zdenek M A L Y (1990) Modified Maeda's scoring system for on-line analysis of CTG. In: Computers in perinatal medicine. K Maeda (ed) Elsevier Science Publishers BV Amsterdam. 239-243
- Caldeyro-Barcia R, Poseiro J J, Negreiros de Paiva C, Gómez-Rogers C, Faúndes-Latham A, Zambrana M A and Arenano-Hernández G. (1963) Effects of abnormal uterine contractions on the human fetus. *Mod Prol Pediat* **8**: 267.
- Caldeyro-Barcia R, Méndez-Bauer C, Poseiro J J, Esarcena L A, Posé S V, Bieniarz J, Arnt I, Gulin L and Althabe O. (1966) Control of human heart rate during labour. In: The heart and circulation in the newborn and infant. Cassels D E.(ed) Grune and Stratton, NY USA. 7-36.
- Caldeyro-Barcia R, deHaan J, Hammacher K, Hon E H, James S, Paul R, Saling E and Sureau C. (1974) Travaux du "Comité d' Amsterdam". *J Gynecol Obstet Biol Reprod.* **3**: 931-934.
- Dawes G S, Visser G H A, Goodman J D S and Redman C W G (1981) Numerical analysis of the human fetal heart rate: the quality of ultrasound records. *Am J Obstet Gynaecol* **141**: 43 - 52
- Dawes G S, Houghton C R S and Redman. (1982) Baseline in human heart-rate records. *Br J Obstet Gynecol.* **89**: 270-275.
- Dawes G S, Moulden M, and Redman C W G (1990) Criteria for the design of fetal heart rate analysis systems. *Int. J Biomed Comput.* **25**: 287-294.
- Dawes G S, Moulden M and Redman C W G (1991a) System 8000: Computerized antenatal FHR analysis. *J Perinat Med* **19**: 47 - 51.
- Dawes G S, Rosevear S K, Pello L C, Moulden M and Redman C W G. (1991b) Computerized analysis of episodic changes in fetal heart rate variation in early labour. *Am J Obstet Gynecol.* **165**: 618-624.
- Divon M Y, Torres F P, Yeh S and Paul R H. (1985) Autocorrelation techniques in fetal monitoring. *Am J Obstet Gynecol.* **151**: 1, 2
- Donker D K. (1991) Interobserver variation in the assessment of fetal heart rate recordings. VU University Press, VU Boekhandel/Uitgeverij bv, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.
- Freeman R. (1990) Intrapartum fetal monitoring - A disappointing story. *N Eng J Med* **322**: 624-626.

- Gabert A J and Stenchever M A. (1976) Continuous electronic monitoring of the fetal heart rate during labour. *Am J Obstet Gynecol* **115**: 919.
- Gagnon R, Campbell M K, Hunse C. (1993) A comparison between visual and computer analysis of antepartum fetal heart rate tracings. *Am J Obstet Gynecol*. **168**: 842-847.
- Grant A. (1992a) EFM alone vs intermittent auscultation in labour. In: Oxford database of perinatal trials. Version 1.3, disk issue 8. Chalmers I. (ed). Oxford: Oxford University Press; Record 3298.
- Grant A. (1992b) EFM + scalp sampling vs intermittent auscultation in labour. In: Oxford database of perinatal trials. Version 1.3, disk issue 8. Chalmers I. (ed). Oxford: Oxford University Press; Record 3297.
- Hammacher K. (1967) Diagnosis of fetal distress by means of cardiotocography. In: proceedings sixth world congress on Gynaecology and Obstetrics. Sydney Australia.
- Hammacher K, Hüter K A, Bokelmann J and Werners P H. (1968) Foetal heart frequency and perinatal condition of the foetus and newborn. *Gynaecologia* **166**: 349-360.
- Hammacher K. (1969) The clinical significance of cardiotocography. Proceedings 1st European congress on perinatal medicine. Huntingford P J, Hunter K A H and Saling, E. (eds) Georg Thieme Verlag, Stuttgart, Germany. 80-93.
- Haverkamp A D, Thompson H E, McFee J G and Cetrulo C. (1976) The evaluation of continuous fetal heart rate monitoring in high risk pregnancy. *Am J Obstet Gynecol* **125**: 310-320.
- Haverkamp A D, Orleans M and Langendoerfer S. (1979) A controlled trial of the differential effects of intrapartum fetal monitoring. *Am J Obstet Gynecol* **134**: 339-408.
- Henson G L, Dawes G S and Redman C W G. (1983) Antenatal fetal heart-rate variability in relation to fetal acid-base status at caesarean section. *Br J Obstet Gynaecol*. **90**: 516-521.
- Hon E H. (1958) The electronic evaluation of the electronic fetal heart rate. Preliminary report. *Am J Obstet Gynecol* **75**: 1215-1230.
- Hon E H. (1959) Observations on 'pathologic' fetal bradycardia. *Am J Obstet Gynecol*. **77**: 1084-1099.
- Hon E H. (1963) The classification of fetal heart rate: I. A working classification. *Obstet & Gynecol*. **22**: 137-146.
- Hon E H. (1967) Detection of fetal distress. In: Fifth world congress in gynaecology and obstetrics. Wood C. (ed) Butterworth & co. Melbourne Australia.
- Hon E H and Quilligan E J. (1967) The classification of fetal heart rate II. A revised working classification. *Conn Med J* **31**: 779-784.
- Hon E H. (1968) An atlas of fetal heart rate patterns. Harty Press, New Haven CT USA
- Jarvis S N, Holloway J S and Hey E N. (1985) Increase in cerebral palsy in normal birthweight babies. *Arch Dis Child* **60**: 1113-1121.
- Kelly V C and Kulkarni D. (1973) Experiences with fetal monitoring in a community hospital. *Obstet. Gynaecol*. **4**: 818-24.

- Kjellmer I (1988) Prenatal and intrapartum asphyxia. In: Fetal and Neonatal Neurology and Neurosurgery. Levene M I, Bennett M J and Punt J (eds) London:Churchill Livingstone 357-369.
- Krause W. (1990) A computer aided monitoring system for supervision of labour. In: Computers in perinatal medicine. Maeda K (ed) Elsevier Science Publishers BV Amsterdam 103-111.
- Langendoer S, Haverkamp A D, Murphy J. (1980) Paediatric follow-up of a randomized controlled trial of intrapartum fetal monitoring. *J Pediatr* 97: 103-107.
- Lawson G W, Belchar R, Dawes G S and Redman C W G (1983) A comparison of ultrasound (with autocorrelation) and direct electrocardiogram fetal heart rate detector systems. *Am J Obstet Gynecol.* 147: 6, 721- 722.
- Leveno K J, Cunningham F G, Nelson S, Roark M, Williams L, Guzick D, Dowling S, Rosenfeld C R and Buckley A. (1986) A prospective comparison of selective and universal electronic fetal monitoring in 34,995 pregnancies. *N Eng J Med* 315: 615-619.
- Lotgering F K, Wallenburg H C S and Schouten H J A. Interobserver and intraobserver variation in the assessment of antepartum cardiotocograms. *Am J Obstet Gynecol.* 144: 701-705.
- Luthy D A, Shy K K, van Belle G, Larson E B, Hughes J P, Benedetti T P, Brown Z A, Effer S, King J F and Stenchever M A. (1987) A randomized trial of electronic fetal monitoring in preterm labor. *Obstet Gynecol* 69: 687-695.
- MacDonald D, Grant A, Sheriden-Pereira M, Boylan P and Chalmers I. (1985) The Dublin randomized controlled trial of intrapartum fetal heart rate monitoring. *Am J Obstet Gynecol* 152: 524-539.
- Maeda K, Kimura S and Fukui Y. (1969) Pathophysiology of the fetus. Fukuoka: Fukuoka Printing, Japan.
- Maeda K, Arima T, Tatsumura M and Nagasawa T. (1980) Computer aided fetal heart rate analysis and automatic fetal distress diagnosis during labour and pregnancy utilizing external technique in fetal monitoring. In: Proc. MEDINFO '80. Lindberg D A B and Kaihara S (eds) North, Holland. 1214-1218.
- Maeda K. (1990) Computerized analysis of cardiotocograms and fetal movements. In: Bailliere's Clinical Obstetrics and Gynaecology. 4: 4, 797-813.
- Neilson J. (1993) Cardiotocography during labour, an unsatisfactory technique but nothing better yet. *BMJ* 306: 347-348.
- Parer J T. (1979) Fetal heart rate monitoring. *Lancet* 2: 632-633.
- Paul R H and Hon E H. (1974) Clinical fetal monitoring V. Effect on perinatal outcome. *Am J Obstet Gynecol* 118: 529-33.
- Rooth G, Huch A and Huch R. (1987) Guidelines for the use of fetal monitoring. FIGO news. *Int J Gynaecol Obstet.* 25: 159-167.
- Schifrin B S and Dame L. (1972) Fetal heart patterns and prediction of Apgar score. *J A M A* 219: 1322.

- Trimbos J B and Keirse M J N C. (1978) Observer variability in assessment of antepartum cardiotocograms. *Br J Obstet Gynaecol.* **94**: 72-75.
- Vincent C A, Martin T and Ennis, M. (1991) Obstetric accidents: the patient's perspective. *Br J Obstet Gynaecol* **98**: 390-395.
- Wood C, Renou P, Oats J, Farrell E, Beisher N and Anderson I. (1981) A controlled trial of fetal heart rate monitoring in a low risk obstetric population *Am J Obstet Gynecol* **141**: 527-534.
- Zuspan F P, Quilligan E J, Iams J D and VanGeijn H P. (1979) Predictors of intrapartum fetal distress: the role of electronic fetal monitoring. *Am J Obstet Gynecol.* **135**: 287-291.
- Johnson J W C, Richards D S, Wagaman R A. (1990) The case for routine umbilical blood acid-base studies at delivery. *Am J Obstet Gynecol.* **162**: 621-625.
- Westgate J, Harris M, Curnow J S H and Greene K R. (1993) Plymouth randomised trial of cardiotocography only verses ST waveform plus cardiotocography for intrapartum monitoring in 2400 cases. *Am J Obstet Gynecol* (In press).
- Westgate J. (1993) An evaluation of electronic fetal monitoring with clinical validation of ST waveform analysis during labour. **MD Thesis**, University of Plymouth Devon UK.

Chapter 2

Optimising the acquisition of the fetal electrocardiogram.

2.1	Introduction	33
2.2	The physical characteristics of fetal scalp electrodes	33
2.3	The clinical performance of fetal scalp electrodes	34
2.4	Quantitative analysis of the signals obtained by fetal scalp electrodes	35
2.5	Signal analysis	38
2.6	Measurement of the electrode frequency responses	44
2.7	Quantitative effects of the electrode frequency responses on the electrocardiogram	46
2.8	Spectral analysis of the fetal electrocardiogram	50
2.9	Discussion and recommendations	54
	Chapter 2 summary	57
	Chapter 2 references	58

2.1 Introduction.

The fetal electrocardiogram (ECG) is obtained from an electrode (fetal scalp electrode, FSE) attached to the fetal scalp or bottom when the fetus is in the breech position. The quality of the fetal ECG waveform obtained by the electrode is inconsistent and is often poor which can hamper monitoring. A study was undertaken to examine the performance of five FSEs and to investigate the nature of the fetal ECG they obtained. This will identify ways in which the acquisition of the signal could be improved (Westgate et al, 1990; Keith et al, 1990). The five electrodes considered in the investigation were; the single spiral Corometrics (Corometrics Medical Systems Inc. USA), single spiral Cetro (Cetro AB, Sweden), the double spiral Hewlett-Packard (Hewlett-Packard Medical, UK) and the Copeland reusable and Copeland disposable (Surgicraft Ltd, UK) electrodes.

The objectives for the study were to,

1. Examine the physical characteristics of the electrodes.
2. Undertake a randomised trial to assess the clinical performance of the electrodes.
3. Analyse the signals obtained by the electrodes.
4. Examine the affect of the electrodes on the ECG.

2.2 The physical characteristics of fetal scalp electrodes.

The term fetal scalp electrode, which perhaps implies a single electrode, is a little confusing because these devices actually comprise two electrodes. The first electrode pierces the fetal presenting part to obtain the fetal ECG which is measured differentially with respect to the second electrode, also located in the head of the device, which makes contact with the vagina via cervical secretions and amniotic fluid. An alternative configuration which is thought to obtain an ECG insensitive to fetal head rotation as the baby moves down the birth canal, has also been suggested (Lindecrantz et al, 1988). Here the fetal ECG is measured differentially with respect to an additional electrode attached to the mother's thigh. The former method is used routinely to obtain an ECG suitable for heart rate calculation.

There are two types of FSEs; the spiral types which have a helical or double helical piercing electrode, and the Copeland types which have a semicircular hook (figure 2.1).

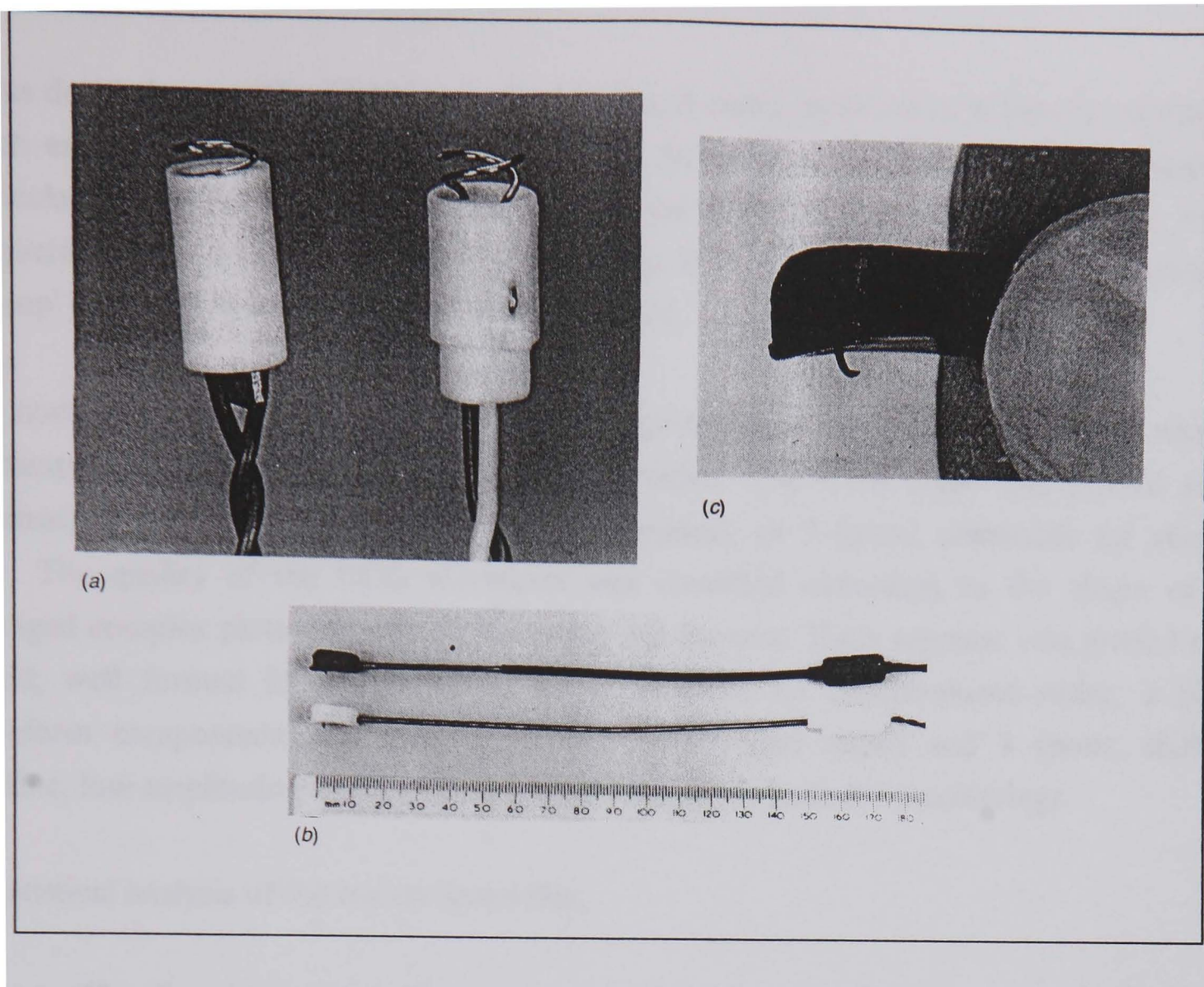


Figure 2.1: Spiral and Copeland FSEs.

(a) Single (left) and double (right) spiral FSEs. (b) Copeland reusable (lower) and disposable (upper) FSEs. (c) Copeland FSE head showing the semicircular piercing electrode held in a half open position.

2.3 The clinical performance of fetal scalp electrodes.

A clinician investigated the clinical performance of the FSEs in a randomised trial. Ten of each of the five electrode types were applied randomly to 50 women in labour after the decision to monitor the patient had been taken by labour ward staff and informed patient consent had been obtained. Only new electrodes of each type were used. All spiral and most Copeland electrodes were applied by the clinician and the remainder were applied by experienced midwives.

The fetal ECG was processed by a ST Analyser (STAN), Cinventa AB, Sweden. This system is unlike those currently used to monitor labour as in addition to recording the CTG, the STAN also provides on-line analysis of the ST segment of the fetal ECG waveform (Rosen and Lindecrantz, 1989). The STAN filters the raw ECG signal for ST-waveform analysis between 0.05 Hz (1st order high pass, passive filter) and 100 Hz (4th order active, Butterworth low pass filter), and for heart rate analysis between 4 and 26 Hz. The signal is digitised to 8 bits at a sampling frequency of 500 Hz. These specifications were confirmed prior to the start of the study.

It was decided to use the STAN recorder because it could interface to a 'lap-top' computer which enabled the digitised ECG to be stored. In addition, as ST waveform analysis is particularly sensitive to signal quality, it allowed the clinician to better identify poor, adequate and good periods of ECG recording. In 20 of the 50 cases chosen at random, a 'lap-top' computer was interfaced to a single STAN.

The complete STAN CTG and ECG complex printouts, were divided up into 30 minute segments and subjectively assessed by the clinician. The CTG trace was graded as 1 (optimal), 2 (fair; adequate for monitoring purposes), or 3 (poor; unsuitable for clinical use). The quality of the ECG waveform was classified according to the shape of an averaged complex plotted by the STAN every 2-3 minutes. Each segment was graded as 1 (good; well formed ECG complexes, stable baseline, no superimposed noise), 2 (fair; waveform components well defined, some superimposed noise) and 3 (poor; shifting baseline, low amplitude signal, noise superimposed, inadequate for monitoring).

A statistical analysis of the results found that,

1. The Copeland disposable electrodes obtained a significantly ($\alpha = 0.05$) lower proportion of grade 1 CTG epochs compared with the Hewlett-Packard, Cetro and Corometrics electrodes.
2. Both Copeland electrode types obtained a significantly ($\alpha = 0.05$) lower proportion of grade 1 ECG complexes and a significantly higher proportion of grade 3 complexes than either the Hewlett-Packard, Cetro or Corometrics electrodes.

2.4 Quantitative analysis of the signals obtained by fetal scalp electrodes.

A study was undertaken to investigate the principal factors which affect the quality of the signals obtained by fetal scalp electrodes. This study extended the clinical investigation and examined the 20 stored raw ECG signals. The objectives were to,

1. Identify the types of noise in the signals.
2. Examine the signal characteristics obtained during grade 1, 2 and 3 periods of ECG recording.
3. Compare the signals obtained by each electrode type.

The signals obtained from the Corometrics and Cetro single spiral electrodes were merged to form a single group termed 'single spiral electrodes' because of their similar construction and clinical performance.

2.4.1 Identification of noise types.

A visual review of the 20 stored records identified the various types of noise and interference which were apparent in the ECG signal.

1. Baseline Shift.

This type of interference has been well documented (Greene, 1987; Kirk and Smith, 1986). If the ECG is only required to obtain a measure of heart rate which requires the detection of the R-wave, then this noise can be removed with suitable bandpass filtering to obtain a stable baseline (figure 2.2a). However, if the ECG is required to analyse the lower frequency variables within the waveform (i.e. the ST-waveform) then removing the baseline is difficult (Froning et al, 1987) as the minimum ECG frequency component of interest is 0.05 Hz (American Heart Association, 1975). This lower limit prevents the filtering of low frequency noise which results in a moving baseline, termed baseline shift on which the ECG is superimposed. This can have amplitudes many times that of the ECG signal (figure 2.2b) and when extreme, can cause saturation in the data acquisition system (figure 2.2c). The first component to be affected will usually be the analogue to digital converter (ADC). For a successful conversion to take place, the analogue signal voltage is required to remain within a certain range (0 - 5 Volts for the STAN). If the analogue signal is amplified above this upper range then the ADC will saturate and the ECG will be lost. If the baseline shift becomes severe it can saturate the amplifiers and active analogue filters, which can then take a considerable time to recover (up to 30 seconds) because of their large capacitance's.

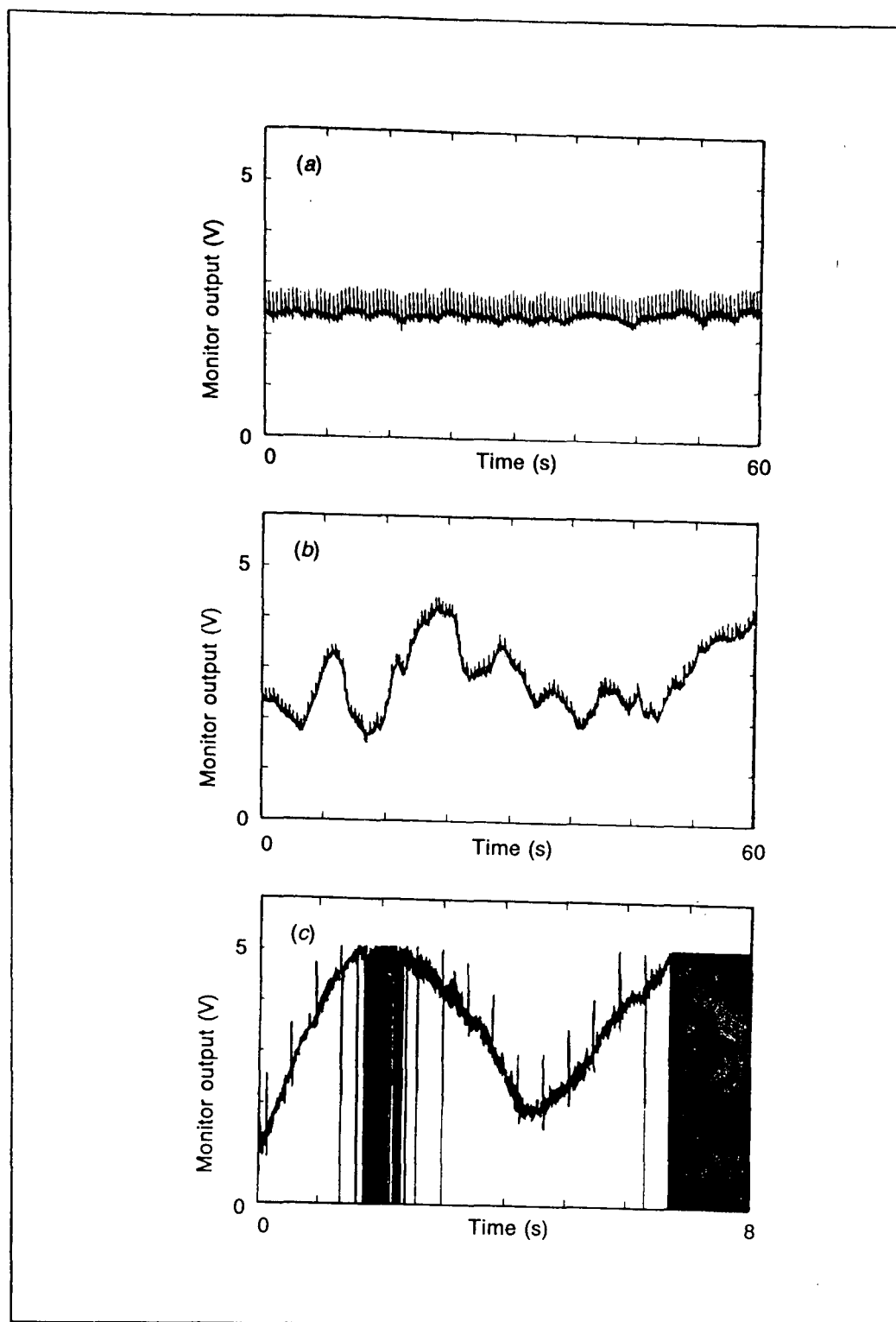


Figure 2.3: Baseline shift and ADC saturation.

(a) A stable baseline, (b) large baseline shifts, (c) Signal saturation caused by severe baseline shifts.

2. Mains interference.

There are two methods for removing mains interference; notch filtering (analogue or digital) in the frequency domain (Van Alste and Schilder, 1985; Proakis and Manolakis, 1988) and adaptive filtering in the time domain (Widrow et al, 1975). The STAN used adaptive filtering prior to data storage (Rosen and Lindecrantz, 1989) and so only low levels of mains interference were expected to be found in the saved data.

3. Signal drop-out

This final parameter was identified during the visual review. It was observed that the ECG was sometimes lost and the signal would drop to some low, non-zero value which it would maintain for up to 20 seconds (10,000 samples).

2.5 Signal analysis.

An analysis of 12 of the 20 stored data records was undertaken. These comprised 3 recordings for each electrode group (single spirals, Hewlett-Packard double spiral, Copeland reusable and Copeland disposable) to examine the presence of the artefacts and their effect upon the quality of the signal. The selected records were considered by the clinician to contain a fair representation of the signal quality obtained by the different electrode groups.

2.5.1 Filter design.

It was proposed to develop three finite impulse response (FIR) digital filters to obtain the baseline frequencies, the mains frequency and the QRS wave frequencies from which the power of each signal could be measured. These measures could then be combined to obtain the QRS signal to baseline noise ratio and the QRS signal to mains noise ratio.

Filter specifications.

Baseline filter: Pass band 0 to 0.5 Hz with transition width, $\Delta f = 1.5$ Hz.

Mains filter: Pass band 49 to 51 Hz with transition width, $\Delta f = 2$ Hz.

QRS wave filter: Pass band 4 to 45 Hz with transition width, $\Delta f = 2$ Hz.

All filters were required to have a maximum pass band ripple of, $A_p = 0.5$ dB and a minimum attenuation of, $A_s = 30$ dB.

In digital signal processing, the number of coefficients, N , required to realise a digital FIR filter is approximately given by equation 2.1 (Bellanger, 1989).

$$N \approx \frac{2}{3} \log_{10} \left(\frac{1}{10 \delta_p \delta_s} \right) \frac{F_s}{\Delta f} \quad (2.1)$$

Where,

F_s is the sampling frequency.

Δf is the transition width between pass band and stop band.

δ_p is given by $A_p = 20 \log_{10}(1 + \delta_p)$ which for $A_p = 0.5$ dB = 0.05925

δ_s is given by $A_s = 20 \log_{10}(\delta_s)$ which for $A_s = -30$ dB = 0.0316

The filters were required to have a narrow transition width, Δf , in the order of 2 Hz, to ensure that only the frequencies of interest were obtained from each filter. Using equation 2.1, with a sampling frequency, $F_s = 500$ Hz, the number of filter coefficients, N required to obtain the required filter response is approximately 289 which was considered impractical. To overcome this the sampling frequency was reduced by a signal processing technique called decimation (Crochiere and Rabiner, 1979).

Decimation of sampling frequency.

The process of decimation reduces the sampling frequency of a signal, $x(n)$, by an integer factor, M , where $x(n)$ has a sampling period T_1 and a sampling frequency $F_1 = 1/T_1$. The decimated signal, $y(n)$, will have a sampling period T_2 and sampling frequency $F_2 = 1/T_2$, which can be expressed as equation 2.2.

$$\frac{T_2}{T_1} = M = \frac{F_2}{F_1} \quad (2.2)$$

$x(n)$ is a full band signal, that is, its spectrum is non-zero for all frequencies in the range $f \leq F_1/2$. To reduce the sampling frequency without incurring aliasing errors, it is first necessary to filter the signal, $x(n)$, with a low pass filter that approximates the ideal characteristic,

$$H(\omega) = \begin{cases} 1, & \omega < \pi/M \\ 0, & \text{otherwise} \end{cases}$$

The sampling rate reduction is then achieved by forming the sequence, $y(m)$ by saving only every M th sample of the filtered output.

The ECG sampling frequency, $F_s = 500$ Hz, was reduced by a decimation factor of $M = 4$, to 125 Hz which meant the desired ECG filter responses could be obtained using a practical number of coefficients ($N < 100$).

The total power, P , was obtained for each extracted signal using equation 2.3,

$$P = 10 \log_{10} \frac{1}{n} \sum_{i=0}^{n-1} (x(i) - \mu)^2 \quad (dB) \quad (2.3)$$

Where, $x(i)$ is the i th sample, n is the total number of samples, and μ is the mean of the samples.

Calculation of the QRS Wave Voltage.

The QRS wave voltage also gives a good indication of signal quality. Ideally the QRS voltage, and therefore the ECG, should be large to ensure it can be accurately resolved. However, this is limited by the baseline shift which causes saturation. A high quality signal would be characterised by a high QRS wave amplitude with low baseline shift. A poor quality signal would be characterised by a low QRS wave voltage with high baseline shift.

The root mean squared voltage (rms) of the QRS wave at the electrode was calculated from the output of the QRS wave extraction filter using equation 2.4.

$$QRS_{rms} = \frac{5}{255.G} \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (x(i) - \mu)^2} \quad (Volts) \quad (2.4)$$

Where, 5/255 is the voltage represented by one least significant bit (5 volts full scale at 8 bit resolution), n is the number of data points, x(i) is the ith data point and μ is the mean of the samples. G is the overall voltage gain of the STAN which could be manually set by clinical staff to 1024, 2048, 4096 and 8192. The gain had a default setting of 2048 and all changes in the setting were automatically saved in the data file.

2.5.2 Signal saturation.

It was found that signal saturation was characterised by the sample values occupying the range $0 < x_i < 15$, or, $240 < x_i < 254$ (254 is the maximum allowed sample value by STAN) with rapid switching between the ranges (figure 2.3(c)). An algorithm was developed to identify these periods lasting longer than 50 samples (0.1 seconds). A period of saturation was considered over when 250 consecutive samples in the unsaturated range were detected.

2.5.3 Signal drop-out.

Signal drop-out was identified when the signal maintained a constant non-zero value in the non-saturated range ($15 \leq x_i \leq 240$) for greater than 100 samples (0.2 seconds).

2.5.4 Software implementation.

The algorithms were implemented using the C-programming language. A flow diagram of the developed software is shown in figure 2.4 and the source code is presented in appendix

A. The software analysed the fetal ECG data in 30 second epochs (15,000 samples). The data was displayed and the amount of saturation and signal drop-out were measured. If both were 0, then the software continued the analysis by decimating the sampling frequency from 500 Hz to 125 Hz. The mains frequency, baseline and QRS waves were then extracted and the signal to noise ratios and the QRS wave rms voltage were calculated and saved to disk.

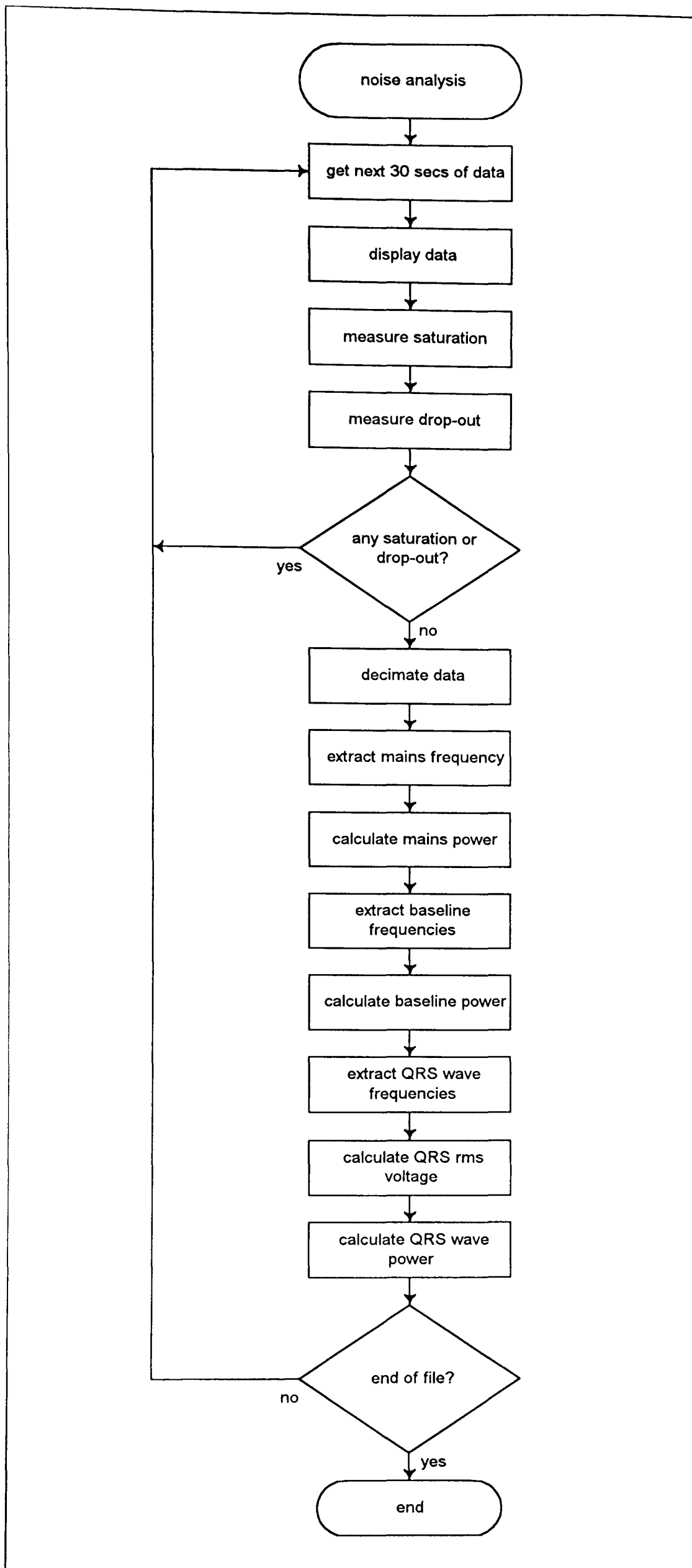


Figure 2.4: Flow diagram of noise analysis software.

2.5.5 Results analysed according to the clinical grading.

During the clinical study, the clinician had identified the periods in the 12 data records for which the ECG was considered, grade 1 (good), grade 2 (adequate) and grade 3 (poor). For each grading, the quantitative results from the analysed epochs were averaged irrespective of the electrode type. This was to investigate the differences in the analysed parameters of the signals between the different classifications. The results for the data records are shown in table 2.1 where μ is the mean and σ^2 is the variance of the given parameter for each grading.

	Grade 1		Grade 2		Grade 3	
	$\mu(1)$	$\sigma^2(1)$	$\mu(2)$	$\sigma^2(2)$	$\mu(3)$	$\sigma^2(3)$
QRS / baseline power ratio (dB)	-9.60	3.08	-14.6	4.43	-16.30	4.13
QRS / mains noise power (dB)	14.04	3.94	9.12	4.89	10.04	5.42
QRS wave amplitude ($\mu\text{V rms}$)	51.14	27.48	43.76	27.63	29.12	20.82
Saturation (average secs/min)	2.35	3.20	8.17	7.40	16.80	12.75
Drop-out (average secs/min)	0.75	0.95	1.07	1.53	1.64	1.96

Table 2.1: Results grouped according to ECG grading

A statistical analysis of these results was applied using the one way analysis of variance test (ANOVA) which was available in commercial software called, 'Minitab' V.7, that ran on a Prime mainframe computer. The null hypothesis was; the signals obtained for each ECG grading were the same, i.e., $\mu(1) = \mu(2) = \mu(3)$. This was tested at the, $\alpha = 0.01$, level of significance.

It was found that,

A period of grade 1 ECG recording had significantly,

1. Higher QRS/baseline signal to noise ratio than grade 2 and grade 3 periods.
2. Higher QRS/mains signal to noise ratio than either grade 2 or grade 3 periods.
3. Greater QRS wave amplitude than grade 3 periods.
4. Less signal saturation than grade 2 and grade 3 periods.

A period of grade 2 ECG recording had significantly,

1. Lower QRS/baseline signal to noise ratio than grade 1 periods.
2. Lower QRS/mains signal to noise ratio than grade 1 periods.
3. Less signal saturation than grade 3 periods.

A period of grade 3 recording had significantly,

1. Lower QRS/baseline signal to noise ratio than grade 1 periods.
2. Lower QRS/mains signal to noise ratio than grade 1 periods.
3. Smaller QRS wave voltage than grade 1 periods.
4. Greater signal saturation than either grade 1 or grade 2 periods.

2.5.6 Results analysed according to electrode type.

The results from the analysed ECG data were also grouped according to the electrode type, irrespective of the grading given by the clinician. This was to investigate which electrodes obtained the best quality signals in terms of the signal parameters being considered. The means, μ and standard deviations, σ , of each parameter measured for each electrode type are shown in table 2.2.

	Single spirals		Hewlett-Packard		Copeland reusable		Copeland disposable	
	$\mu(ss)$	$\sigma^2(ss)$	$\mu(hp)$	$\sigma^2(hp)$	$\mu(cr)$	$\sigma^2(cr)$	$\mu(cd)$	$\sigma^2(cd)$
QRS / baseline noise ratio (dB)	-8.22	2.49	-9.21	4.34	-13.72	3.60	-16.68	4.37
QRS / mains noise ratio (dB)	16.16	0.74	14.45	4.64	12.04	4.14	11.05	1.87
QRS wave amplitude (μV rms)	49.51	7.11	67.58	25.20	25.28	6.14	21.31	7.51
Saturation (average secs/min)	3.52	4.99	5.71	7.80	5.75	7.13	14.48	11.36
Drop-out (average secs/min)	0.67	0.72	0.62	0.91	1.59	1.13	0.97	1.08

Table 2.2: Signal analysis results grouped according to electrode type

A one way analysis of variance was applied to these results using a similar method as previously described. The null hypothesis was that the signal parameters measured for each electrode have the same means, $\mu(ss) = \mu(hp) = \mu(cr) = \mu(cd)$. This was tested at the $\alpha = 0.01$ level of significance. It was found that,

Single spiral electrodes have significantly,

1. Higher QRS/baseline signal to noise ratio than all other electrode types.
2. Higher QRS/mains signal to noise ratio than all other electrode types.
3. Greater QRS wave amplitude than both Copeland electrodes.
4. Lower signal saturation than other electrode types.

Hewlett-Packard double spiral electrodes have significantly,

1. Higher QRS/baseline signal to noise ratio than both Copeland types.
2. Higher QRS/mains signal to noise ratio than both Copeland types.
3. Greater QRS wave amplitude than other electrode types.
4. Lower signal saturation than Copeland disposable electrodes.

Copeland reusable electrodes have significantly,

1. Higher QRS/baseline, signal to noise ratio than Copeland disposable electrodes.
2. Higher QRS/mains signal to noise ratios than Copeland disposable electrodes.
3. Greater QRS wave amplitude than Copeland disposable electrodes.
4. Lower signal saturation than Copeland disposable.

Copeland disposable electrodes have significantly,

1. Lower QRS/baseline, signal to noise ratio than other electrode types.
2. Lower QRS/mains signal to noise ratio than other electrode types.
3. Smaller QRS wave amplitude than other electrode types.

2.6 Measurement of the electrode frequency responses.

It has been suggested that the frequency bandwidth of interest for the ECG is, 0.05 - 100 Hz (American Heart Association, 1975). The frequency response of the 5 FSEs were measured over this range in an in-vitro experiment represented in figure 2.5.

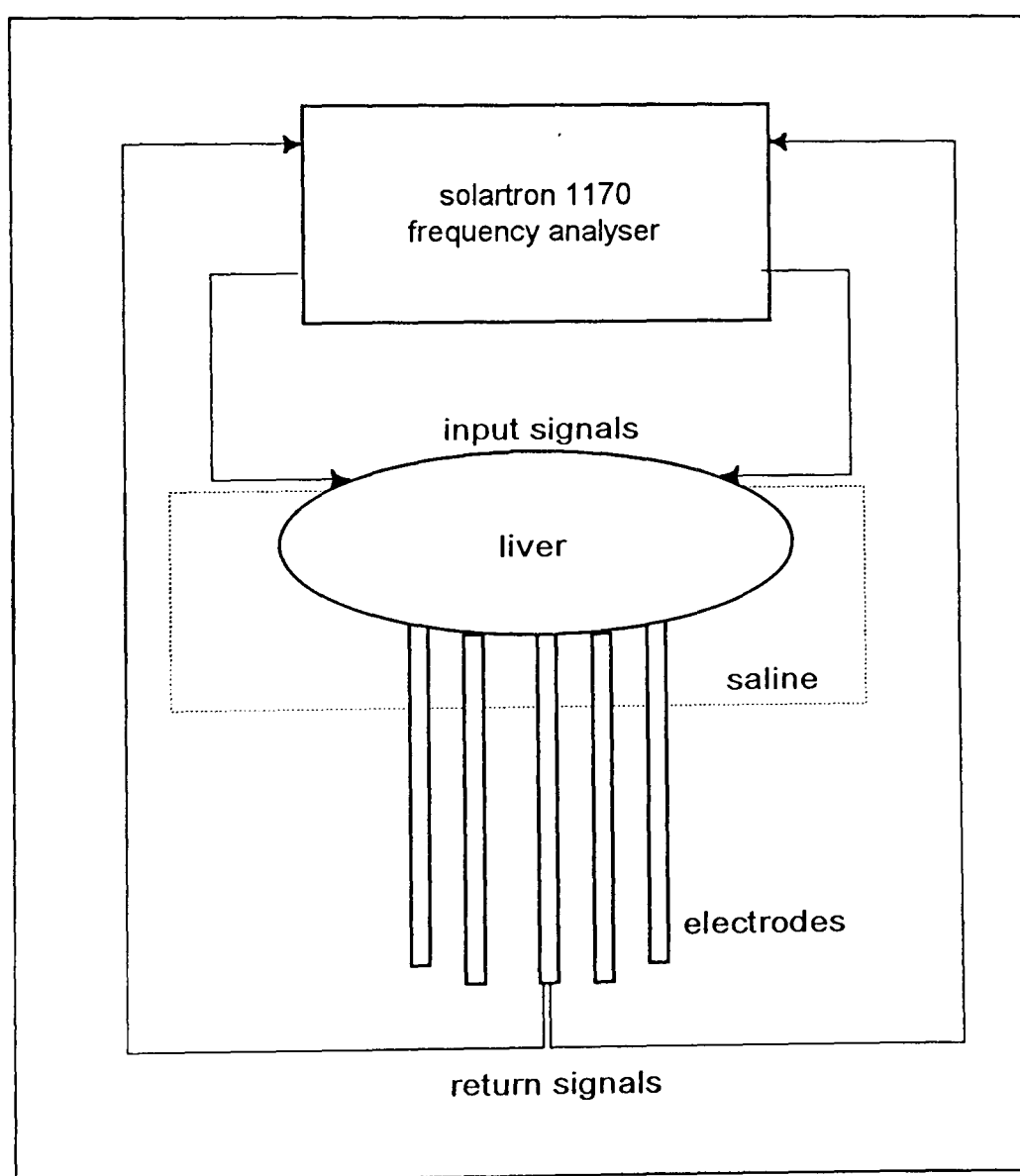


Figure 2.5: Experiment to obtain the electrodes frequency responses.

To simulate the electrode/tissue interface, a fresh pig's liver was immersed to a depth of 1 cm in a saline bath and the electrodes were optimally attached to the lower lobe of the liver by the clinician. Sinusoidal signals (1 Volt pk-pk) from a Solartron 1170 frequency response analyser, with frequencies in the range 0.01 to 200 Hz were introduced into the liver using hypodermic needles. The signal obtained by the electrode under test was then fed back to the Solartron for analysis. The Solartron and CTG recorders have similar input characteristics with near infinite input impedances provided by an active buffering stage.

2.6.1 Results

Over the range 0.01 - 200 Hz (70 frequencies in all) the gain and phase of the signal obtained by each electrode was recorded. A maximum phase change of 14° was measured which was considered negligible (Copeland reusable). The frequency responses obtained are shown in figure 2.6.

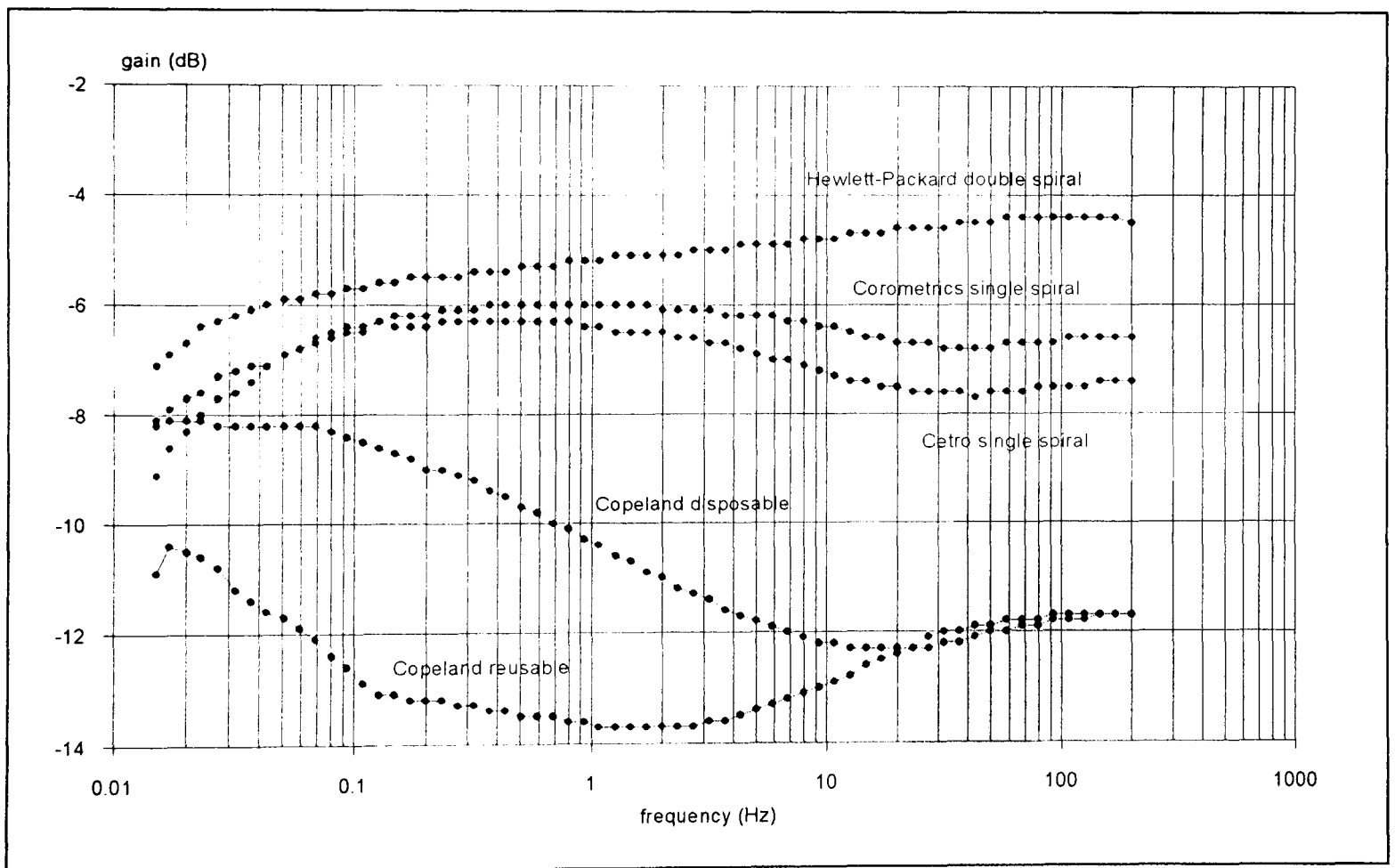


Figure 2.6: The frequency responses of the FSEs.

As expected, the signals from the Solartron analyser were attenuated by the impedance of the liver and fluids. Interestingly, the relative attenuations between the electrodes were different. The Hewlett-Packard double spiral electrode obtained a higher amplitude signal than the single spiral electrodes, which in turn obtained larger signals than either of the Copeland types. The frequency characteristics of the electrodes fell largely into two groups; the responses obtained for the spiral FSEs approached the ideal, being nearly flat within the frequency band of interest (0.05 - 100 Hz). The Copeland FSEs attenuated the whole signal

more than the spirals, and had a non-linear response, which attenuated higher frequency components 4 to 8 dBs more than the very low frequency components. This response is undesirable because it would accentuate the baseline shift frequencies.

2.7 Quantitative effects of the electrode frequency responses on the electrocardiogram.

The implications of the electrodes' frequency response on the ECG shape in the time domain was investigated. The fetal ECG could not be used for this study because it cannot be obtained without a fetal scalp electrode. Consequently, the signal would be modified by the electrode used during data collection. To overcome this, the ECG was collected from four neonates within 1 hour of birth which have an ECG similar to the fetus. Precordial silver/silver-chloride skin electrodes were used which minimally affect the ECG. Two digital filters were designed to have the same frequency response as the Copeland reusable and the Corometrics single spiral electrodes. These electrodes were chosen as they represented the two different frequency response groups observed in figure 2.6. The neonatal ECGs were then passed through the electrode filter models and the outputs compared.

2.7.1 Filter design using the frequency sampling method.

The electrode frequency response models were designed using the frequency sampling method (Proakis and Manolakis, 1988; Lockhart and Cheetham, 1989). This technique requires that the frequency responses be sampled linearly. The accuracy of the model is dependent upon the number of samples taken. The values obtained are regarded as discrete Fourier transform (DFT) coefficients from which the filter coefficients for the desired response can be obtained using the inverse discrete Fourier transform (IDFT).

The frequency samples are obtained by sampling the desired frequency response which is given by $H_d(e^{j\omega})$ at N points, ω_k , for $k = 0, 1, 2, \dots, N-1$, uniformly spaced around the unit circle such that,

$$\omega_k = \frac{2\pi k}{N}$$

Sampling the desired frequency response at these frequencies gives,

$$H(k) = H_d(e^{j\omega})|_{\omega=\omega_k}$$

$$H(k) = H_d(e^{j2\pi k/n})$$

These samples are regarded as DFT samples from which the frequency coefficients, $h(n)$, can be calculated using the IDFT (equation 2.5).

$$h(n) = \frac{1}{N} \sum_{k=0}^{N-1} H(k) e^{2\pi i k n / N} \quad \text{where } n = 0, 1, \dots, N-1. \quad (2.5)$$

2.7.2 Implementation of the frequency sampling method to obtain the electrode frequency response models.

The frequency sampling method required that the electrode frequency responses be sampled linearly with frequency and then normalised to the sampling frequency, F_s which was 1000 Hz. This would mean that some of the lower frequency features within the Copeland reusable response would be normalised to very narrow widths indeed (figure 2.7).

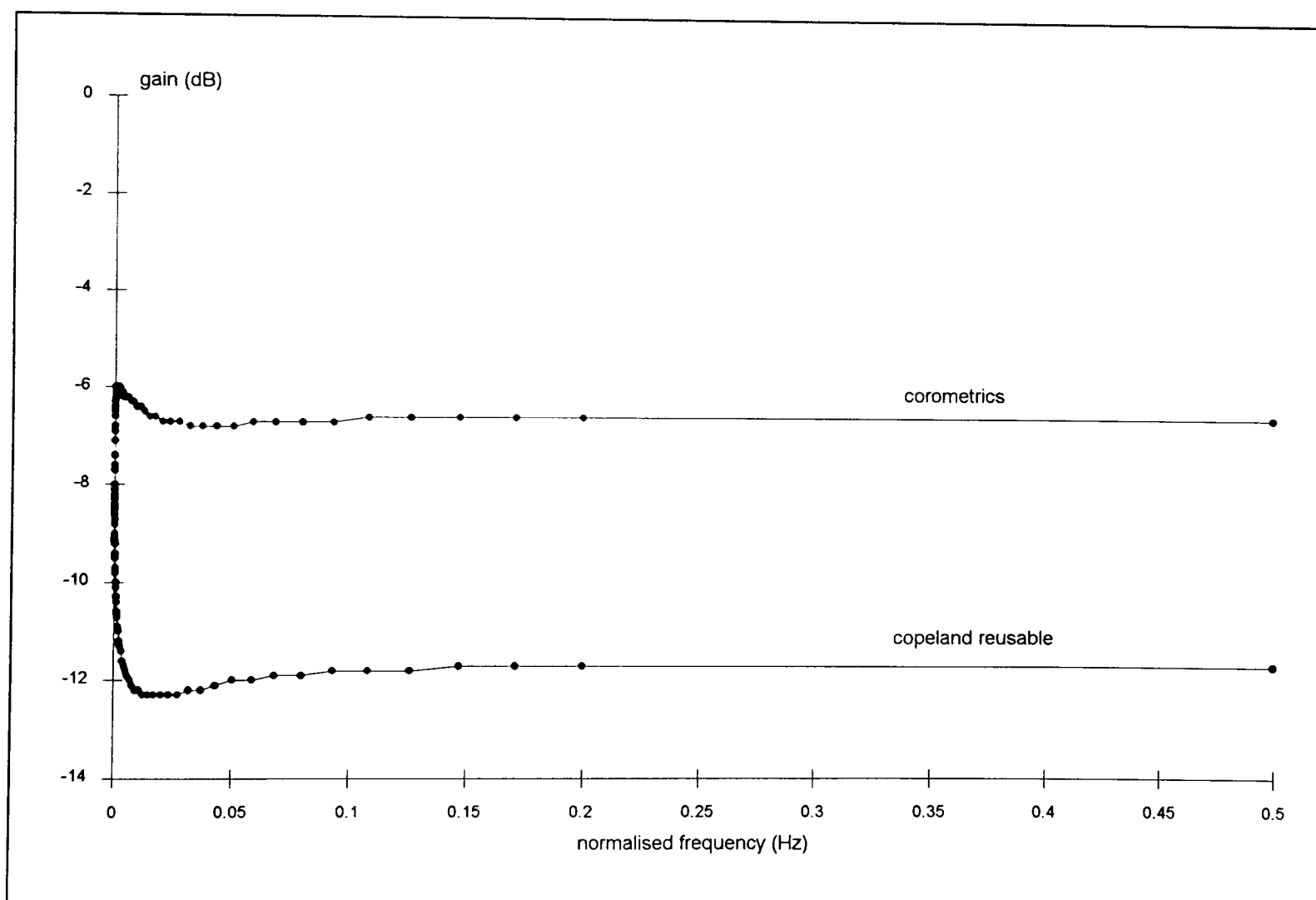


Figure 2.7: Electrode frequency responses on a linear scale normalised to 1000 Hz.

The peaked region between 0.01 Hz and 0.07 Hz, for example would have a normalised width of just 0.00006 Hz, which would require an impractical number of samples to obtain an accurate model. For this reason the sampling frequency of the neonatal ECG data was

decimated by a factor of $M = 4$, from 1000 Hz to 250 Hz using the same procedure as previously described (section 2.4.1). The frequency response curves normalised to the decimated sampling frequency, $F_s = 250$ Hz are shown in figure 2.8.

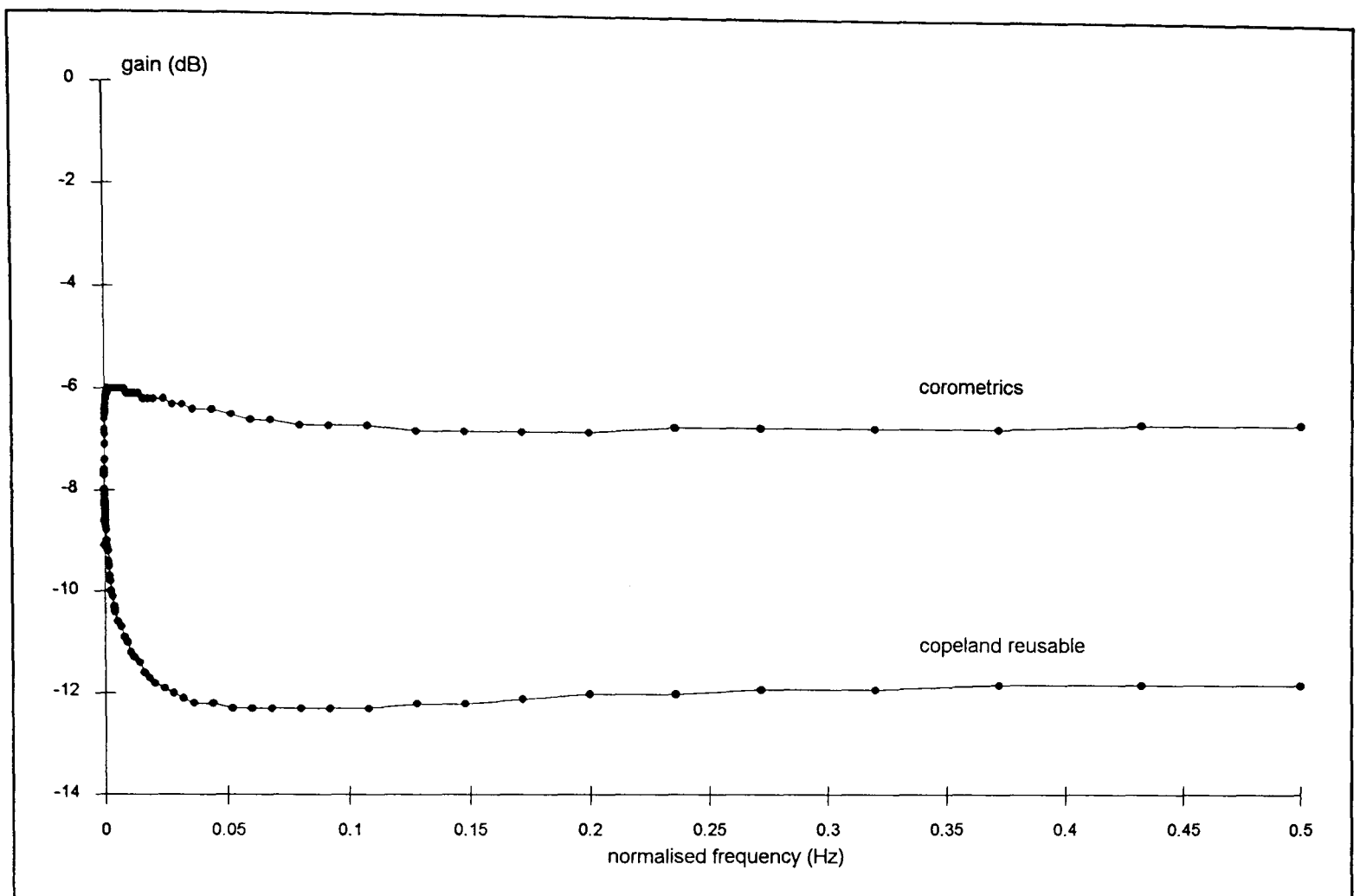


Figure 2.8: Electrode frequency responses normalised to 250 Hz

The frequency responses were sampled at intervals of 0.002 Hz to obtain 251 samples for each electrode. These were then converted back from decibels to obtain 251 DFT coefficients. Software was designed to implement the IDFT to produce 513 filter coefficients which were truncated using the Hamming window. This high number of coefficients was necessary to realise the still narrow features. As a check, the frequency response of the calculated coefficients was obtained using a commercial signal processing package, Interactive Laboratory System (ILS) (Signal Technology Inc. California USA 1987) and is shown in figure 2.9. A visual comparison of modelled frequency responses (figure 2.9) and the actual frequency response (figure 2.8) confirmed that the electrode frequency characteristics had been accurately modelled.

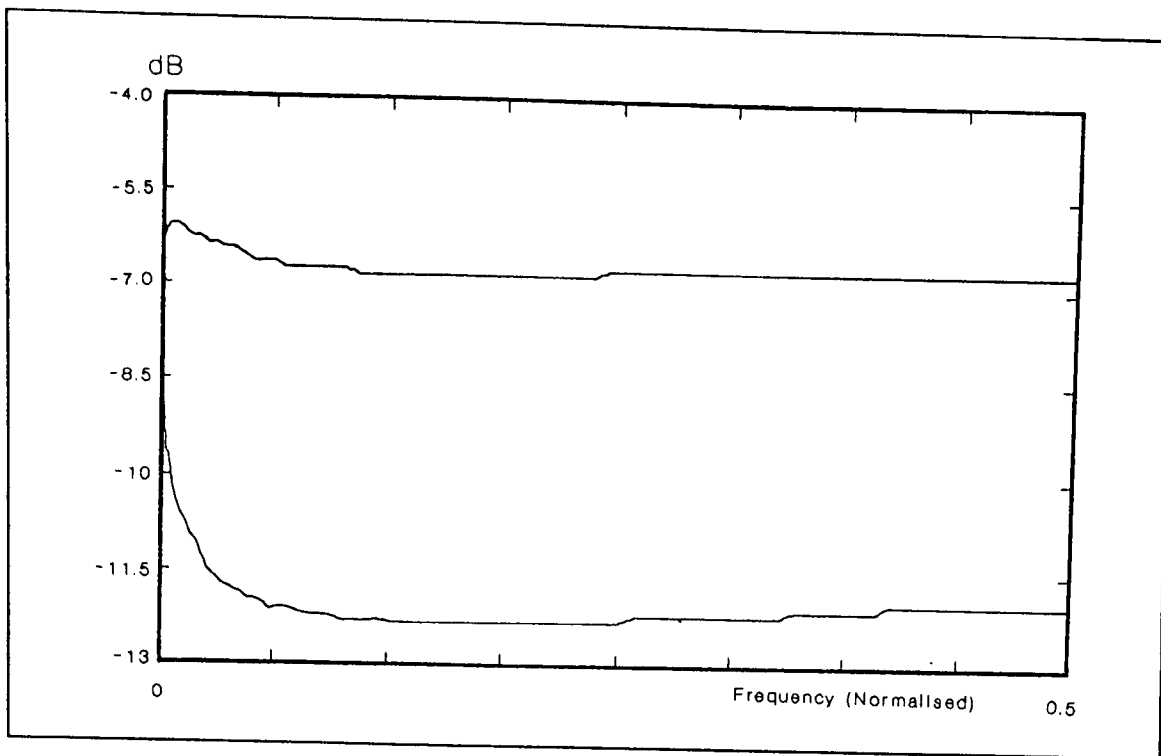


Figure 2.9: Electrode Frequency Response Models

2.7.3 Results.

The derived filter coefficients were convolved with the decimated neonatal ECG data to simulate the effects of the electrodes (figure 2.10).

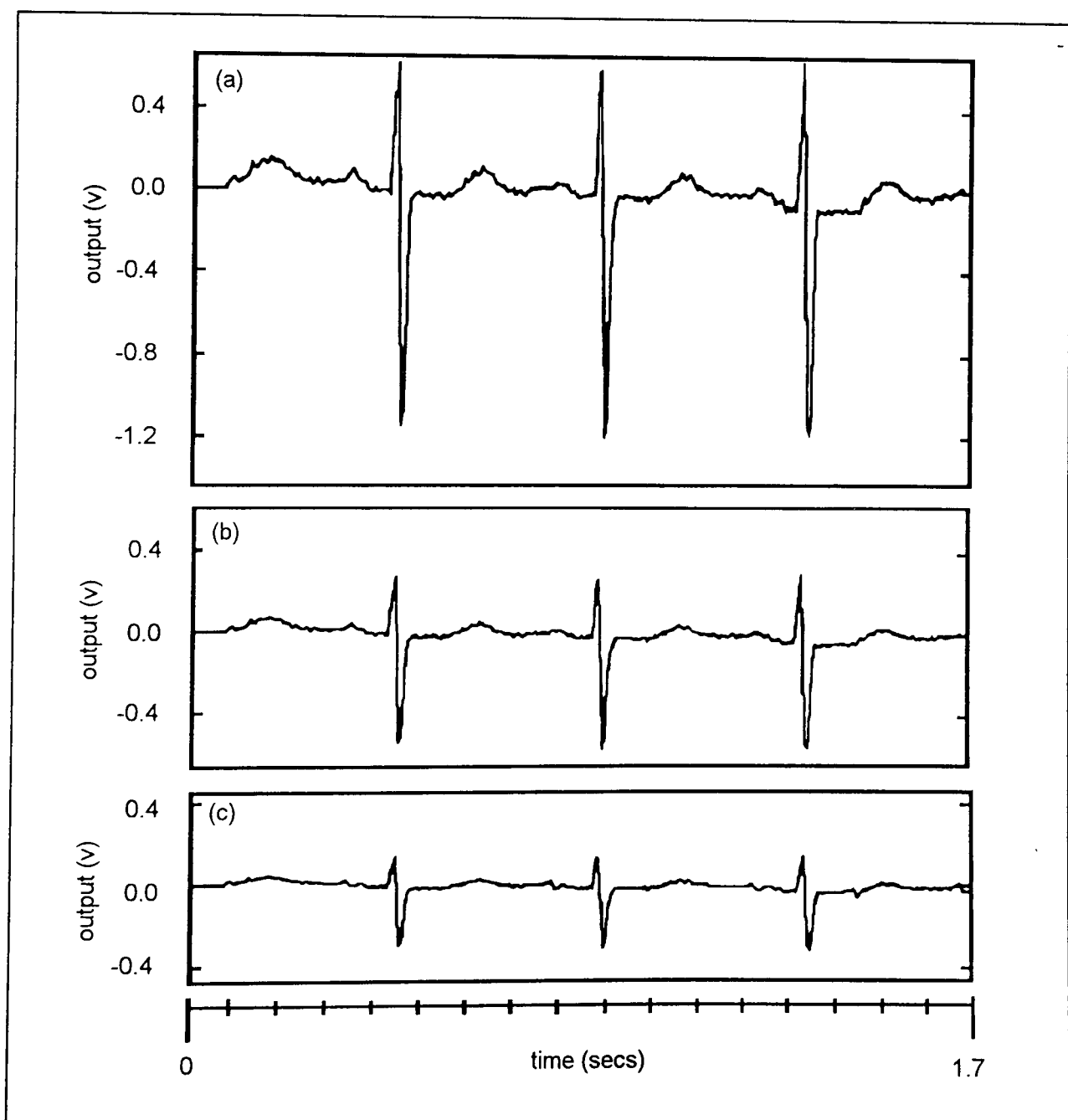


Figure 2.10: Electrode simulation results

Figure 2.10(a) shows a segment of neonatal ECG signal before filtering. Figure 2.10(b) shows the effects of filtering this segment with the Corometrics model and figure 2.10(c) shows the effects after filtering with the Copeland reusable model. As expected, visual examination shows that both electrode models attenuated the signal, which was due to the impedance of the liver and fluids during the original experiment. The results show that the signal filtered with Copeland reusable model was 54% of the peak to peak amplitude of the Corometrics model. While there did not appear to be any major changes in the overall waveform shapes, there did appear to be modifications in the P and T wave regions associated with the Copeland FSE.

2.8 Spectral analysis of the fetal electrocardiogram.

The location of the spectral bands of the QRS, R, P and T waves within the fetal ECG were investigated. This was to examine further the effects of the non-linear electrode frequency responses on the specific regions of the ECG complex.

Six fetal ECG records, obtained using single spiral electrodes, were reviewed. When a period of grade 1 (good) recording was found, software was implemented to decimate the sampling frequency of the data from 500 Hz to 125 Hz. For each record, an averaged ECG complex was obtained using 20 consecutive complexes which improved the signal to noise ratio (Sheild and Kirk, 1981; Kirk and Smith, 1986).

Several conventional methods were considered to obtain the bandwidth of the QRS, R, P and T wave regions within a given averaged ECG complex, but because of the discontinuous nature of the averaged waveform and the small number of data points making up each region, these methods were not suitable. Therefore, a slightly unconventional method was devised; 11 low pass and 10 high pass digital FIR filters were designed.

1. Low pass filters.

The cut-off frequency ranged from 5 to 55 Hz in 5 Hz intervals.

Transition width, $\Delta f = 5$ Hz.

Minimum stop band attenuation, $A_s = 30$ dB.

Maximum pass band ripple, $A_p = 0.5$ dB.

2. High pass filters.

The cut-off frequency ranged from 1 to 10 Hz in 1 Hz intervals.

Transition width, $\Delta f = 1$ Hz.

Minimum stop band attenuation, $A_s = 30$ dB.

Maximum pass band ripple, $A_p = 0.5$ dB.

The six averaged ECG complexes were passed through each filter and the filtered data values obtained from the 21 filters were saved. The un-filtered complexes were then displayed and the QRS, R, P and T waves were visually identified. The data values making up each identified region were then compared to the corresponding data values of each of the filtered regions.

The premise for the following analysis was that there would only be a significant difference between an un-filtered region and corresponding filtered region if the given filter had removed some of the important frequencies; there would be no difference in the values if the filter had been ineffectual. To assess when a significant difference had occurred a statistical method was employed. This made use of the fact that if a filter had had no effect, then if the un-filtered sample values were plotted against the corresponding filtered values, then a straight line would be obtained with gradient, $\beta = 1$ and intercept, $\alpha = 0$ ($y = x$).

The un-filtered data together with each filtered data set in the different regions of interest were considered as paired data-sets. Normal regression analysis assumes that for each fixed sample, $x(i)$ (un-filtered data), the conditional density of the corresponding random variable, $xf(i)$ (filtered data), has a normal distribution. This condition is valid because if a filter had no effect then the only difference between the un-filtered and filtered samples would be random noise effects. For a given paired data set, regression analysis allowed a straight line to be fitted where the estimated gradient, $\hat{\beta}$ is given by equation 2.6.

$$\hat{\beta} = \frac{\sum_{i=1}^N x(i) xf(i) - \frac{1}{N} \sum_{i=1}^N x(i) \sum_{i=1}^N xf(i)}{\sum_{i=1}^N (x(i))^2 - \frac{1}{N} \left(\sum_{i=1}^N x(i) \right)^2} \quad (2.6)$$

$$\text{and where the intercept, } \hat{\alpha} = \overline{xf(i)} - \hat{\beta} \overline{x(i)} \quad (2.7)$$

for $i = 1, 2, \dots, N$

where,

$\hat{\beta}$ is the estimated gradient.

$\hat{\alpha}$ is the estimated intercept.

$x(i)$ is the i th un-filtered data sample.

$xf(i)$ is the i th filtered data sample.

N is the total number of data pairs.

The random variables, t_{α} and t_{β} , can then be generated from equations 2.6 and 2.7 to have

the statistical t-distribution with N-2 degrees of freedom (Freund and Walpole, 1987).

The null hypothesis for a given set of paired data samples was that the filter had no effect ($x(i) = xf(i)$) and when plotted a straight line with gradient, $\beta = 1$, and intercept, $\alpha = 0$, would be obtained. The alternative hypothesis was that the filter had removed some of the important frequencies ($x(i) \neq xf(i)$) and when plotted a straight line with $\beta \neq 1$ or $\alpha \neq 0$ would be obtained. The significance level for this test was 0.05.

To illustrate the procedure, consider figure 2.11.

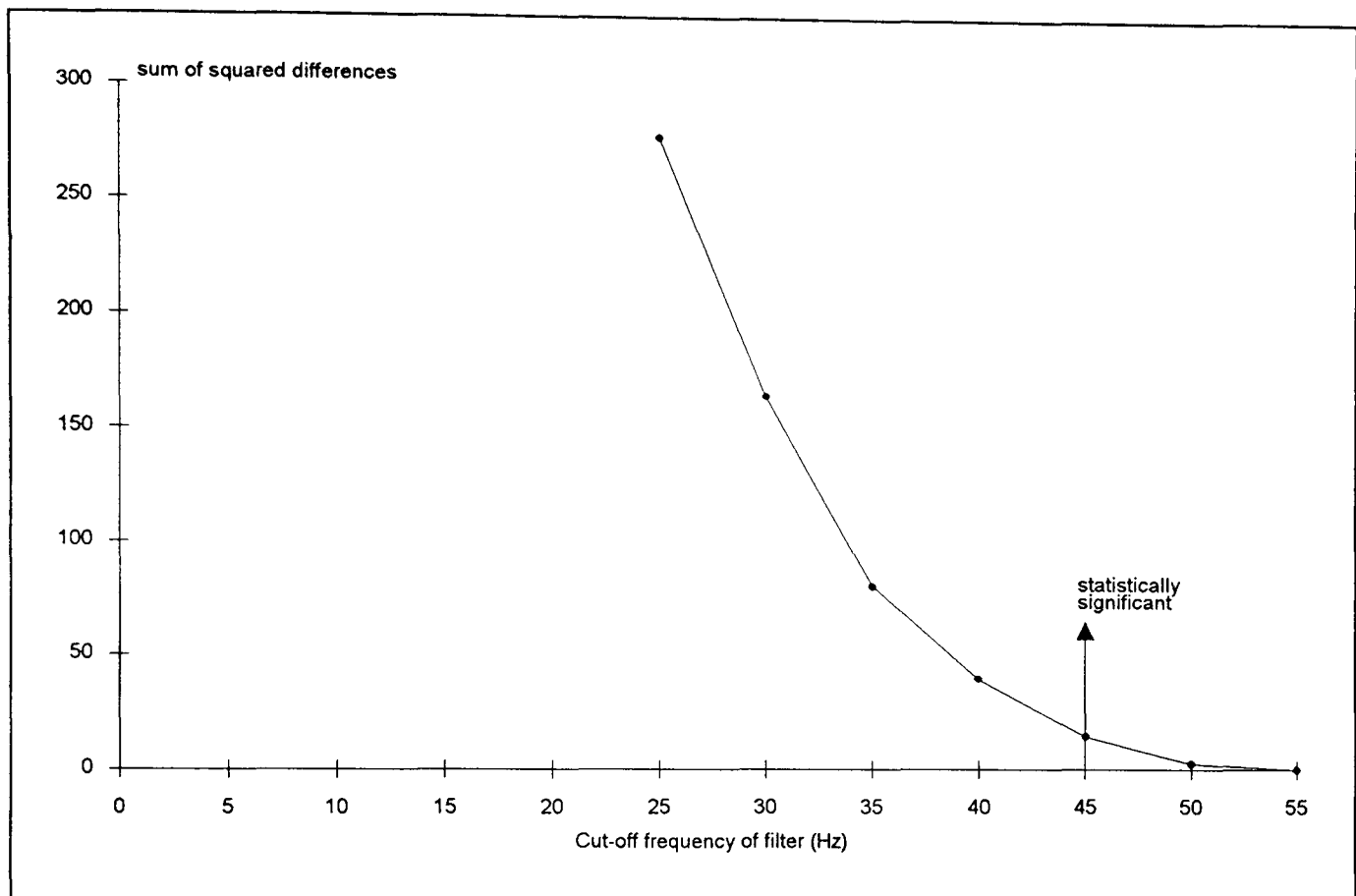


Figure 2.11: Low Pass Filtering Effects on QRS Wave.

Figure 2.11 shows how the sum of the squared differences between an un-filtered and filtered QRS wave increases as the cut-off frequency of the low pass filter is reduced. When the filter cut-off frequency was 45 Hz, the differences between the un-filtered and filtered data were statistically significant. Therefore, the upper frequency of interest for this QRS wave was 50 Hz because this was the highest filter cut-off frequency not to significantly affect the waveform. The lowest frequency of interest was found in a similar way using the high pass filters.

This procedure was carried out for the QRS, R, P and T waves of all 6 averaged ECG complexes. The averaged results for each region are summarised in table 2.3.

	Lower band		Upper band	
	mean (Hz)	standard deviation	mean (Hz)	standard deviation
QRS - wave	1.8	0.9	48.3	2.4
R - wave	4.2	0.7	49.8	2.1
P - wave	0.7	0.4	22.5	4.8
T- wave	-	-	11.7	2.4

Table 2.3: The averaged frequency bands of the QRS, R, P and T waves.

The lower band of the T-wave could not be found because all the high-pass filters used significantly modified this region. The lowest cut-off frequency used was 1 Hz, so the lower T-wave frequency of interest must be below this. The measured frequency bands were then plotted on the electrode frequency response curves to indicate the regions of the fetal ECG likely to be modified by a non-linear electrode frequency response (figure 2.13).

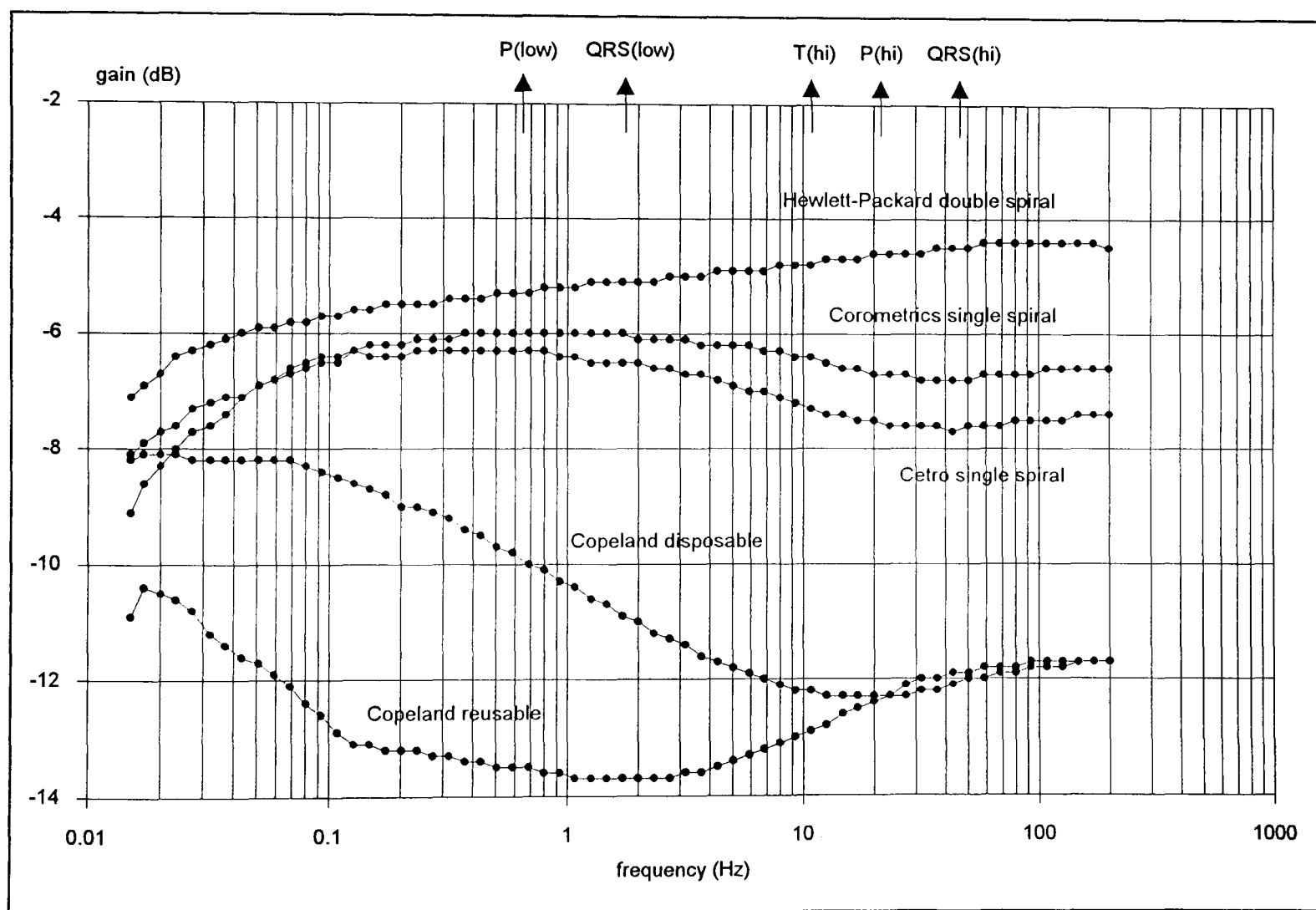


Figure 2.12: Electrode frequency responses with ECG frequency bands indicated.

Figure 2.12 shows that the near flat frequency response of the spiral electrodes are unlikely to significantly modify the ECG. However, the non-linear frequency responses of the Copeland electrodes are likely to modify all regions of the ECG, but particularly the P and T waves.

2.9 Discussion and recommendations.

There has only been one previous reported comparison of spiral and Copeland electrodes (Nickelsen et al 1989). This compared clinically the Copeland disposable and Corometrics FSEs for heart rate monitoring only and found no differences in complication rates which is in agreement with the randomized clinical trial of this study.

The factors which influence the quality of ECG recording have been investigated. It was found that the signal associated with a good period of recording has significantly higher QRS amplitude, lower baseline shift and lower saturation rates than other grades. Conversely, the signals obtained during a poor period of recording have significantly lower QRS wave amplitude, higher baseline shift and higher levels of signal saturation. Interestingly, the only significant difference between an adequate period and a poor period, was the signal saturation caused by very large baseline shifts. This suggests that it was baseline shift (and in the extreme, saturation) which largely determines the quality of fetal ECG recording.

If the ECG is only required for heart rate calculation using the R-wave, then the baseline shift can be removed from an unsaturated signal with suitable bandpass filtering. This study suggests that the R-wave bandwidth is 4 - 50 Hz. A more selective range may be preferred to further reduce baseline shift and mains interference at the expense of smaller R-waves. However, with a flat baseline, these could be subsequently amplified. If the ECG is required for waveform analysis then the frequency range of interest is taken as 0.05 - 100 Hz, consequently baseline shifts cannot easily be removed. This frequency range was recommended by the American Heart Committee some time ago and relates to the specification for a chart recorder required to record the adult ECG which can be obtained with a more stable baseline (American Heart Association, 1975). The validity of this lower frequency for the fetal ECG has never been established. When a fetal ECG complex is extracted for analysis from the continuous signal, it may be that this lower limit is excessive and could be increased as it is difficult to image how signals with a period of 20 seconds (0.05 Hz) could significantly influence an ECG complex which lasts less than half a second except to raise or lower the isometric reference. This would be especially true if successive ECG complexes were averaged as this process has an inherent filtering effect of their own. However, until work is undertaken to establish the minimum bandwidth for the fetal ECG these guidelines should be adhered to.

The difficulty in fetal ECG data acquisition is to suitably amplify the signal without allowing baseline shifts to move outside the ADC window causing signal saturation. The STAN recorder digitised the ECG to 8-bit resolution which meant the sampled data had a range 0 - 255. This range may be too narrow for the ECG to be adequately resolved for waveform analysis without risking saturation. If 12 bit resolution were used (sample range 0 - 4095)

then the ECG could be suitably resolved and the risks of saturation considerable reduced.

This study also examined the suitability of fetal scalp electrodes for obtaining the fetal ECG. The Hewlett-Packard double spiral FSE was found to have a near ideal frequency response and obtained the highest amplitude ECGs. However, in the clinical situation these electrodes are difficult to apply which makes their recording quality inconsistent. This was substantiated by the quantitative analysis which found a large standard deviation associated with the QRS wave collected by this electrode type. The Copeland electrodes are difficult to attach optimally, obtain smaller ECGs and have increased baseline shift. In addition their frequency response was found to be non-linear which accentuates the baseline shift and modifies all regions of the ECG, but in particular the P and T waves. The problems associated with these electrodes may in part be caused by their long semi-rigid design and method of attachment. This can be seen in figure 2.13 which shows an X-ray of a woman in early labour with a Copeland electrode attached to the fetal scalp.

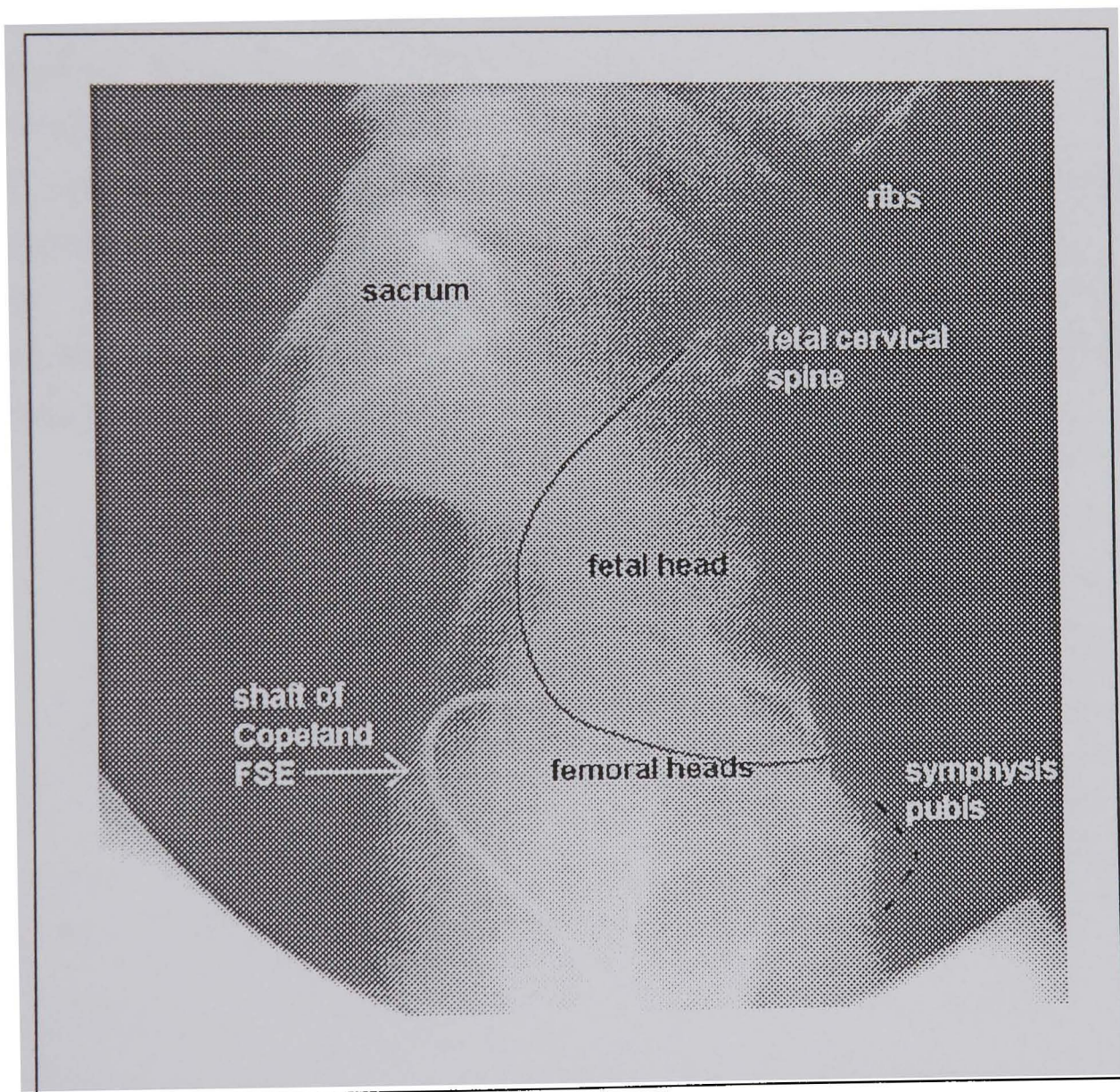


Figure 2.13: X-ray showing the attachment of the Copeland electrode.

Notice that the sprung shaft of the Copeland electrode is flexed which causes the electrode head to move with maternal movements which is likely to increase baseline shift.

The single spiral electrodes obtained an ECG with twice the amplitude of the Copeland types and also had a near ideal frequency response. Clinically, these electrodes are the easiest to apply optimally and provide a secure, consistent attachment to the fetus and are therefore the fetal scalp electrodes considered most suitable for obtaining the fetal ECG.

The fetal ECG is routinely used to calculate heart rate but research shows that there may be additional important information to be obtained from other variables within the waveform (Greene, 1983, 1987; Murray, 1986; Rosen, 1986). The findings of this study have important implications for this research because they suggest the causes of inadequate signal quality can be addressed if care is taken in the design of the data acquisition system for which the minimum recommended requirements are,

1. Single spiral fetal scalp electrode.
2. ADC resolution of 12 bits to ensure adequate waveform resolution with low risk of saturation from excessive baseline shifts.
3. Sampling frequency of 500 Hz.
4. Mains filter: Either adaptive or twin 'T' notch filter.
5. ECG bandwidth for waveform analysis of 0.05 - 100 Hz .
6. ECG bandwidth for heart rate of 4 Hz - 50 Hz maximum. A more selective range may be preferred.
7. Signal quality assessment; if the acquisition system could monitor the ECG then automatic gain control could be used to optimally amplify the digitised signal and minimise saturation.

Chapter 2 summary.

The nature of the fetal electrocardiogram (ECG) as obtained by 5 commonly available fetal scalp electrodes (FSEs) was investigated. These electrodes fall broadly into two groups; spiral electrodes which have a single or double helical spike to pierce the fetal presenting part, and Copelands which use a hook. 10 of each electrode were used to record the ECG from 50 patients in labour of which 20 were digitised and stored. The clinician visually reviewed the paper printouts from the cases and subjectively graded the periods of recording as good, adequate and poor.

The objectives for the study were to,

1. Identify the types of noise which influence the recording quality.
2. Analyse the signals obtained by the electrodes.
3. Examine the affect of the electrodes on the ECG.

The types of noise which were thought to influence the quality of the signal were low frequency baseline shifts, mains interference, signal saturation and signal drop-out. It was found that the principal factor which determined signal quality was baseline shift which, when extreme, caused signal saturation. The frequency of this noise is within the bandwidth of the ECG and so cannot be easily removed. It was found that the spiral electrodes obtained significantly ($\alpha = 0.05$) lower levels of baseline shift and signal saturation than the Copeland types.

The frequency responses of the electrodes were obtained in an in-vitro experiment which simulated the electrode/tissue interface. The spiral electrodes were found to have a frequency response which approached the ideal, being relatively flat in the frequency band of interest. The Copeland electrodes were found to have a non-linear response which attenuated the signals more than the spirals and accentuated baseline shifts. These frequency responses were modelled to investigate the time domain implications. It was found that spiral electrodes obtain signals with twice the amplitude of the Copelands. The frequency bands of the important features of the ECG were obtained and it was found that the non-linear frequency responses of the Copeland electrodes were likely to modify all features, but particularly the lower frequency P and T-waves.

This study suggests that the effects of baseline shift can be reduced with careful consideration to the design of the data acquisition system and by using a single spiral electrode.

Chapter 2 references.

- American Heart Association Committee on Electrocardiography. (1975) Recommendations for the standardization of leads and of specifications for instruments in ECG/VCG Circulation. **52**: 11-25.
- Bellanger M. (1989) Digital processing of signals; theory and practice. John Wiley and Sons Ltd Chichester England. 2nd edition, 128.
- Crochiere R E and Rabiner L R (1979) A program for multistage decimation, interpolation and narrow band filtering in programs for digital signal processing. In: Digital Signal Processing Committee IEEE Acoustics, Speech and Signal Processing Society (eds.) **8.3**: 1-14.
- Freund J E and Wolpole R E. (1987) Mathematical Statistics Prentice-Hall, 4th edition **14.4**: 463-473.
- Froning J N, Froelicher V F and Olsen M D (1987) Applications and limitations of continuous baseline estimation using a cubic-spline technique during exercise ECG testing. *Comp Cardio*. 537-540.
- Greene K R. (1983) Quantification of ST waveform changes of the fetal electrocardiogram and their relationship to asphyxia. MD Thesis, Department of Obstetrics and Reproduction University of Southampton.
- Greene K R. (1987) The ECG waveform. In: Ballière's clinical obstetrics and gynaecology. **1**: 1, 131-155.
- Keith R D F, Ifeachor E C, Curnow J S H, Westgate J and Greene K R. (1990) Quantitative Assessment of the effects of different scalp electrodes on the Fetal Electrocardiogram. 1st International Symposium on Intrapartum Surveillance University Hospital Nottingham England 18-19 October. (abstract only)
- Kirk D L Smith P R. (1986) Techniques for the routine on-line processing of the fetal electrocardiogram. *J Perinat Med*. **14**: 391-397.
- Lindecrantz K, Lilja H, Widmark C and Rosen K G. (1988) Fetal ECG during labour: a suggested standard *J Biomed Eng*. **10**: 351-353.
- Lockhart G B and Cheetham B M G. (1989) Basic Digital Signal Processing. Butterworths Kent England 95-99.
- Murray H G. (1986) The fetal electrocardiogram: current clinical developments in Nottingham. *J Perinat Med*. **14**: 399-404.
- Nickelsen C, Weber T, Parnell C, Nim J, Kemp A M and Junge I. (1989) Cardiotocographic monitoring of deliveries. A prospective investigation of two types of electrodes *Ugeskr Laeg*. **151**: 440-442.
- Proakis J G and Manolakis D G. (1988) Introduction to digital signal processing. Macmillan Publishing Company New York.
- Rosen K G. (1986) Alterations in the fetal electrocardiogram as a sign of fetal asphyxia - experimental data with a clinical implementation. *J Perinat Med*. **14**: 335-363.

- Rosen K G and Lindecrantz K. (1989) STAN - the Gothenburg model for fetal surveillance during labour by ST analysis of the fetal electrocardiogram. *Clin Phys Physiol Meas.* **10**: Suppl B, 51-56.
- Sheild J E A and Kirk D L. (1981) The use of digital filters in enhancing the fetal electrocardiogram. *J Biomed Eng.* **3**: 44-48.
- Van Alste J A and Schilder T S. (1985) Removal of base-line wander and power-line interference from the ECG by an efficient FIR filter with a reduced number of taps. *IEEE Trans Biomed Eng.* **32**: 12, 1052-1060.
- Westgate J, Keith R D F, Curnow J S H, Ifeachor E C and Greene K R. (1990) Suitability of fetal scalp electrodes for monitoring the fetal electrocardiogram during labour. *Clin Phys & Physiol Meas.* **11**: 4, 297-306.
- Widrow B, Glover J R, McCool J M, Kaunitz J, Williams C S, Hearn R H, Zeilder J R, Doug E, and Goodlin R C. (1975) Adaptive noise cancelling principals and applications *Proc. IEEE* **63**: 1692-1716.

Chapter 3

Development of an intelligent system for labour management:

Feature extraction from the cardiogram.

3.1	Introduction	61
3.2	Feature extraction using artificial neural networks	62
3.3	A neural network to classify of the complete cardiogram	71
3.4	A neural network to classify the magnitude and timing of heart rate decelerations	74
3.5	A neural network to classify the magnitude of decelerations	76
3.6	Discussion of the suitability of neural networks for feature extraction .	81
3.7	Baseline heart rate estimation	82
3.8	Heart rate accelerations	87
3.9	Heart rate variability	89
3.10	Contractions	92
3.11	Signal quality	96
	Chapter 3 summary	98
	Chapter 3 references	99

3.1 Introduction.

The conceptual model from which the intelligent system for labour management was based is shown in figure 3.1.

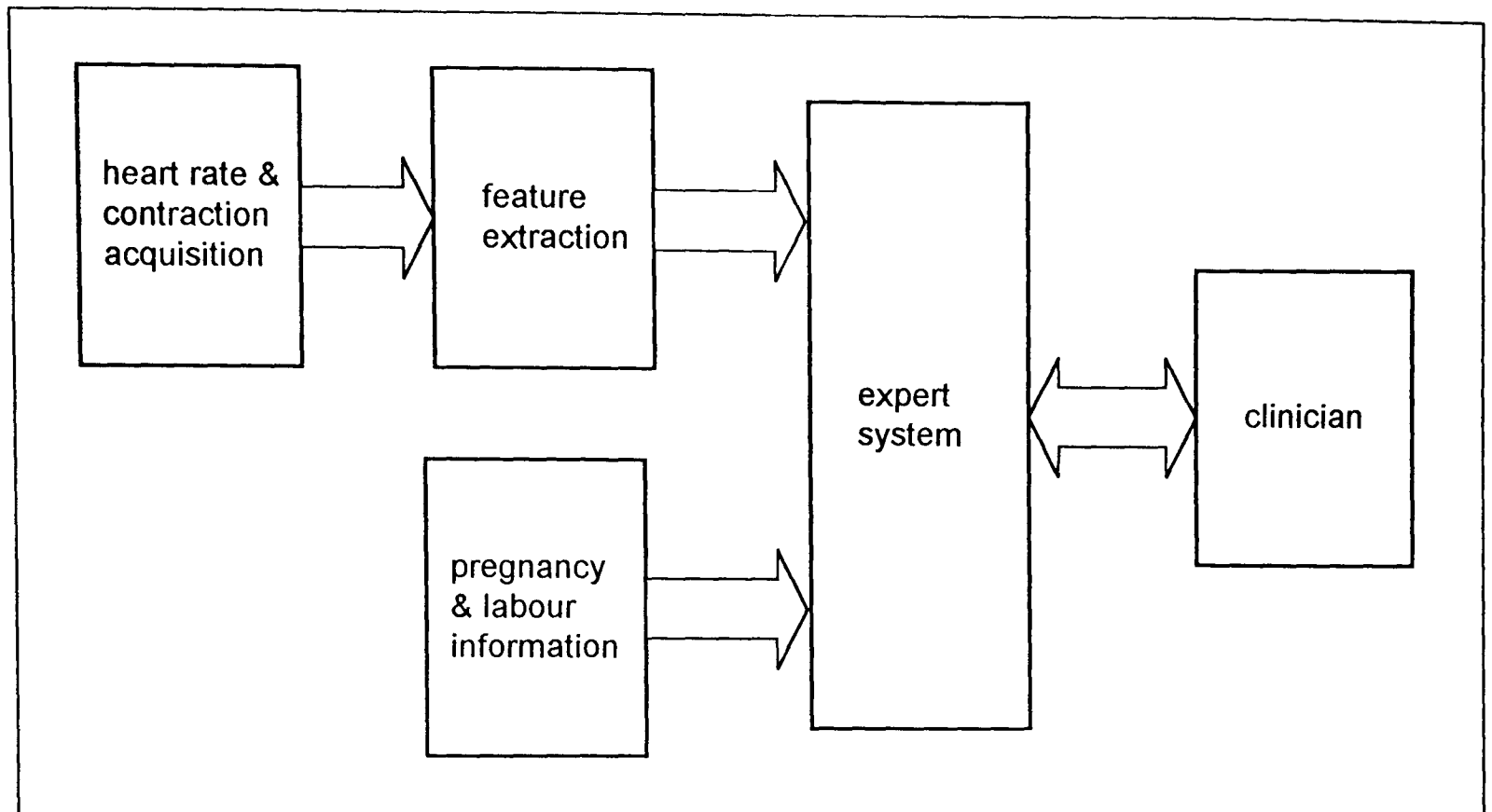


Figure 3.1: Conceptual model of an intelligent system for labour management.

It was proposed to obtain the fetal heart rate and uterine contraction signals from a conventional intrapartum CTG recorder. These signals would pass through a process of feature extraction to obtain the important information. This, together with the information pertaining to the specific pregnancy and labour would then be processed by an expert system.

The approach adopted throughout this investigation was to attempt to produce a system which extracted and processed information in the same way and at the same level of performance as the expert obstetrician. Experts can be difficult to identify especially in an area such as obstetrics where the clinicians' performance is not monitored. The criteria used to identify the obstetric expert were therefore subjective. However it was considered important to apply two conditions for accepting an individual as an expert.

1. The clinician must be experienced and still practising on the labour ward.
2. The clinician should be regarded as 'expert' by their peers and superiors.

In addition, the success of the project would require that experts be open, articulate, motivated and be aware of, and want the possible benefits an intelligent system could bring

to the labour ward. It was also felt that the success of the project would depend on the contributions of more than one expert.

The development of the system was undertaken along two main paths; the methods for feature extraction from the CTG and the expert system, which were later merged to form the complete system. This chapter considers the methods developed for feature extraction.

The features of the cardiotocogram.

The features within the CTG which are thought to reflect fetal condition have been described in detail in chapter one. The key features are:

1. Baseline heart rate; about which the heart rate pattern fluctuates.
2. Heart rate variability; the amplitude of high frequency perturbations about the baseline.
3. Accelerations in heart rate; relatively long term transient increases in heart rate from the baseline.
4. Decelerations in heart rate; relatively long term transient decreases in heart rate from the baseline, classified according to their size and timing in relation to uterine contractions.
5. Frequency and timing of contractions.

An additional important parameter is a measure of the quality of the heart rate and contraction signals.

3.2 Feature extraction using artificial neural networks.

It was previously recognised that the performance of the system would depend on its ability to extract and classify features from the CTG in the same way and at the same level of performance as experts. Previous attempts to automate the feature extraction process have used empirical methods (Krause, 1990; Maeda, 1990; Dawes et al, 1991). Detailed studies have not been published to demonstrate how closely these methods compare with the experienced clinician. In contrast, artificial neural networks (NNs) are trained by experts using examples representative of the problem and so do not depend on empirically derived means. This may allow them to perform at a level comparable with experts. It was for these reasons that it was decided to investigate how NNs might be implemented to extract and

classify complex features from the CTG and compare their performance with experts and conventional algorithms (Keith et al, 1993).

3.2.1 The backpropagation neural network.

The backpropagation NN model was used for these investigations and is shown in figure 3.2. This NN is a feed-forward, fully connected network with one input layer, (h), one hidden layer of nodes, (i), and an output layer of classification nodes, (j). All investigations were performed on an 80486 microprocessor based personal computer. The software was developed in the C-programming language and is given in appendix B.

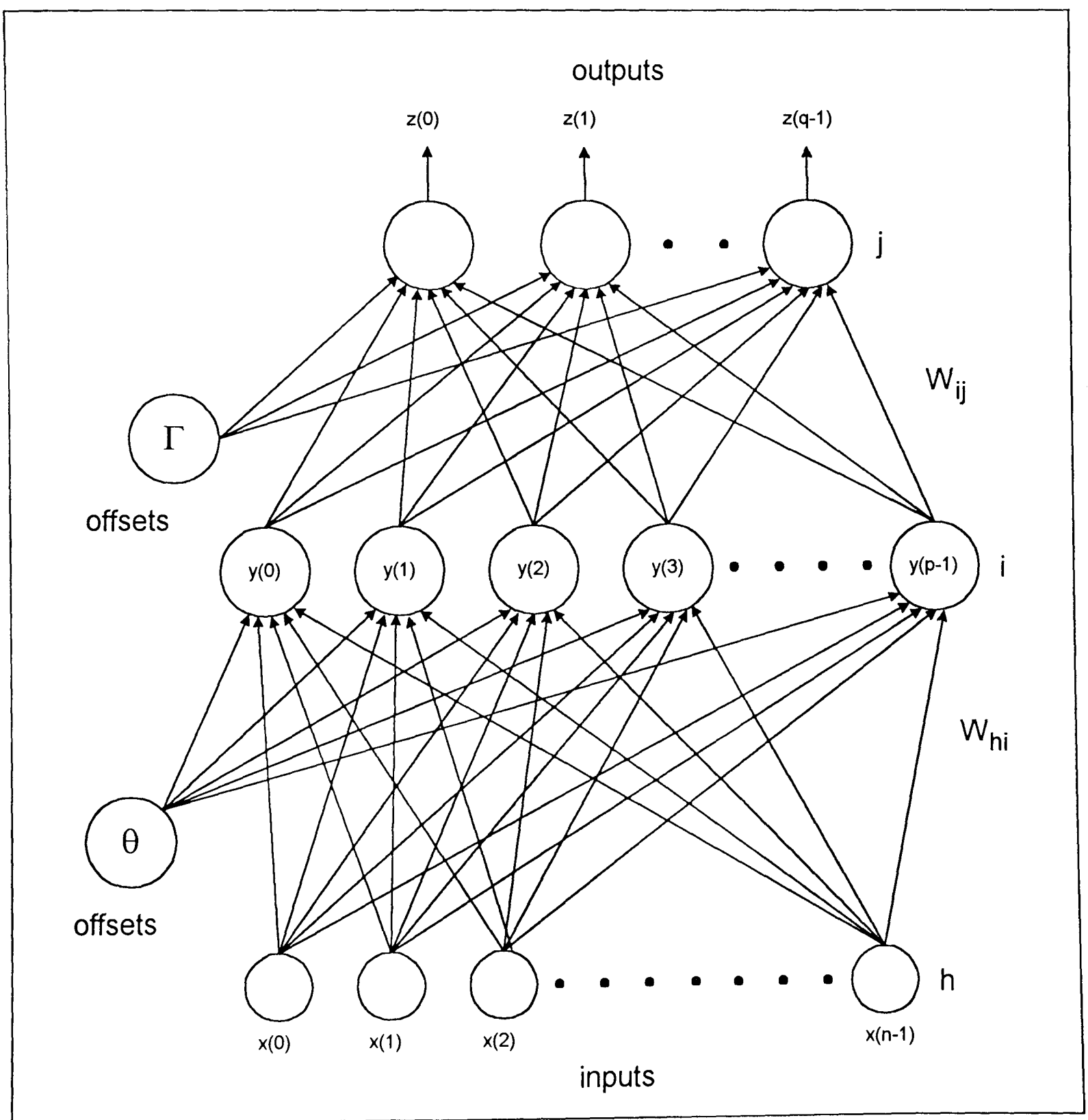


Figure 3.2: The backpropagation neural network model.

General description of operation.

The input matrix, $X(h) = [x(0), x(1), x(2) \dots x(n-1)]$, is formed from the sampled input data to be classified and is presented to the input layer, h , of the NN. Each node of the hidden layer, i , receives a signal associated with each input sample, modified by the interconnecting weight, W_{hi} . The signals received at each node are summed and an offset, θ_{hi} , is added. A threshold function, $f(x)$, is applied to the sum total to produce a nodal output, $y(i)$, in the range 0 to 1 (figure 3.3).

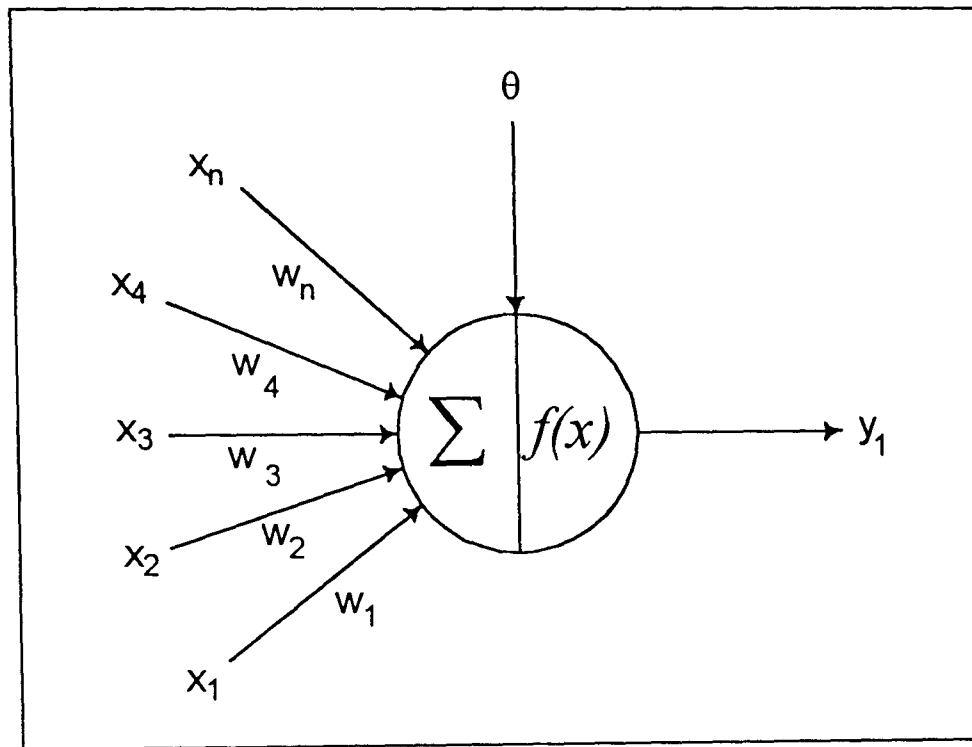


Figure 3.3: A processing element or node.

The operation of the node is described mathematically by equation 3.1.

$$y = f\left(\sum_{i=1}^n w_i x_i - \theta\right) \quad (3.1)$$

Where the thresholding function, $f(x)$, is usually taken as the sigmoid function described by equation 3.2.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3.2)$$

which is shown graphically in figure 3.4.

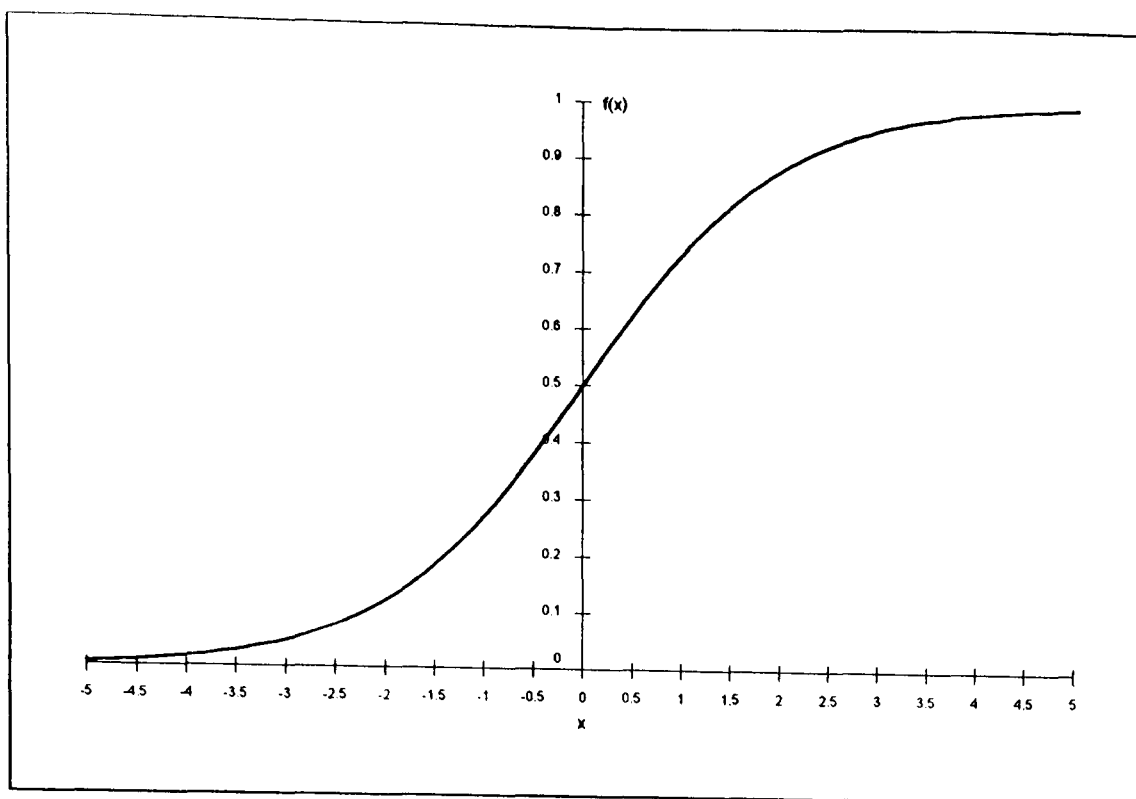


Figure 3.4: Sigmoid threshold function.

The hidden layer matrix, $Y(i) = [y(0), y(1), y(2) \dots y(p-1)]$, is processed in a similar way through the next layer of weights, W_{ij} , to form the output matrix, $Z(j) = [z(0), z(1), z(2), \dots z(q-1)]$. The NN is trained to generate a specific output pattern when a specific feature is present in the input data.

Training the backpropagation neural network.

Artificial neural networks, unlike conventional approaches, are trained rather than programmed. Programming would take the form of attempting to understand and formalise the expert knowledge into a set of procedures and rules which could then be directly implemented by a computer.

A backpropagation NN is trained using a comprehensive library of examples compiled and classified by a domain expert. The interconnection weights, W_{hi} and W_{ij} , are firstly randomised to signed non-zero values. A training example is selected from the library at random and processed through the NN. A measure of the error between the actual output pattern achieved and the desired output pattern is calculated. A small change proportional to the error and specified by the backpropagation learning algorithm (Rumelhart et al, 1986), is made to the weights to bias the network towards the desired output response. The process of selecting an example and biasing the weights is repeated many times until the NN approaches a global solution where the output error for a high proportion of the examples is small and lies within a previously defined limit. At this point the NN is said to have converged and the learning process is complete. The performance of the NN can then be assessed by comparing its classification of 'unseen' examples with experts. The backpropagation algorithm is now described.

The backpropagation algorithm.

1. Random values are assigned to all the interconnection weights, W_{hi} and W_{ij} , and to each of the offsets, θ_i and Γ_j .
2. The input pattern matrix, $X(h) = [x(0), x(1), x(2) \dots x(n-1)]$, and the desired output pattern matrix, $D(j) = [d(0), d(1), \dots d(q-1)]$, form the pattern pair which are presented to the network. For pattern classification, the outputs are generally made binary.
3. The input pattern matrix, $X(h)$, is filtered through the interconnection weights, W_{hi} , using equation 3.3, for all, p , middle layer nodes.

$$y(i) = f\left(\sum_{h=0}^{n-1} x(h)w_{hi} + \theta_i\right) \quad (3.3)$$

Where, $y(i)$ is the activation value of the i th middle layer node, θ_i is the i th middle layer offset value, and $f(x)$ is the logistic sigmoid threshold function described in equation 3.2.

4. The middle layer output matrix, $Y(i)$, is filtered through the interconnection weights, W_{ij} , using equation 3.4, for all q output nodes.

$$z(j) = f\left(\sum_{i=0}^{p-1} y(i)w_{ij} + \Gamma_j\right) \quad (3.4)$$

Where $z(j)$ is the activation value of the j th output node and Γ_j is the j th output node offset value.

5. The error between the computed output and the output desired by the expert is found with equation 3.5, for all q output nodes.

$$\varepsilon_j = z(j)(1 - z(j))(d(j) - z(j)) \quad (3.5)$$

Where d_j is the output for node j desired by the expert and ε_j is the computed error of the j th output node.

6. The error for each middle node output, relative to each ε_j is calculated using equation 3.6, for all p middle nodes.

$$\varepsilon_i = y(j)(1 - y(j)) \sum_{j=0}^{q-1} w_{ij} \varepsilon_j \quad (3.6)$$

Where q is the number of output nodes and ε_j is the computed error associated with the i th middle node.

7. Each interconnection weight between the middle layer and output layer are modified by equation 3.7,

$$w_{ij}(t+1) = w_{ij}(t) + \alpha \varepsilon_j y(i) + \mu(w_{ij}(t) - w_{ij}(t-1)) \quad (3.7)$$

Where the learning rate gain term, α , is in the range, $0 < \alpha < 1$. This parameter prevents the network from oscillating due to large adjustments in the weights and thereby helps the network slowly converge to a general solution for all input patterns in the training set. The momentum term, μ , is used to speed up the learning process by adding a proportion of the previous weight adjustment to the present change. The rationale for this term takes the following form. For each interconnection weight, the desired weight solution lies in a certain particular direction from random starting position. The continued presentation of input data leads the value of the weight to be modified with a velocity. By adding a proportion of the previous weight change, a given weight is effectively accelerated towards its final solution thus improving the learning time.

8. The offset values for the output layer are adjusted by equation 3.8, for all q outputs.

$$\Gamma_j(t+1) = \Gamma_j(t) + \alpha \varepsilon_j + \mu(\Gamma_j(t) - \Gamma_j(t-1)) \quad (3.8)$$

9. The weights of the interconnections between the inputs and middle layer are adjusted by equation 3.9, for all p middle nodes.

$$w_{hi}(t+1) = w_{hi}(t) + \alpha \varepsilon_i x_h + \mu(w_{hi}(t) - w_{hi}(t-1)) \quad (3.9)$$

10. The middle layer offsets are adjusted with equation 3.10, for all p middle nodes.

$$\theta_j(t+1) = \theta_j(t) + \alpha \varepsilon_j + \mu(\theta_j(t) - \theta_j(t-1)) \quad (3.10)$$

11. The process is then repeated from step 2 until the global error of the network becomes sufficiently low.

Once training is complete the network weights are saved. The NN can then be used to classify 'unseen' data by forming the input matrix, $X(h)$, and processing this through the saved weights, W_{hi} and W_{ij} , using equations 3.3 and 3.4, to form the output classification matrix $Z(j)$.

An observation regarding the backpropagation learning algorithm.

In steps 5 and 6, an assessment of the error in the NN output was obtained using equations 3.5 and 3.6, which took the form,

$$\varepsilon_j = z(j)(1 - z(j))(d(j) - z(j))$$

These equations have some properties which on the face of it appear rather strange. The roots of the equation which determines the error, ε_j , are given when an output, $z(j)$, is 0 or 1, or when the output equals the desired output, $d(j)$. This last root seems reasonable. However, the roots when, $z(j) = 0$ and $z(j) = 1$ are independent of the desired output. This suggests that when an output is actually 0, but is desired to be 1, the error is 0. Similarly, when $z(j) = 1$ and $d(j) = 0$. An explanation of this anomaly could not be found and so it was investigated. Consider the graph of $y = z(1 - z)$ (with roots $z = 0$ and $z = 1$) for the range $0 < z < 1$, which corresponds to the range of the NN output shown in figure 3.5.

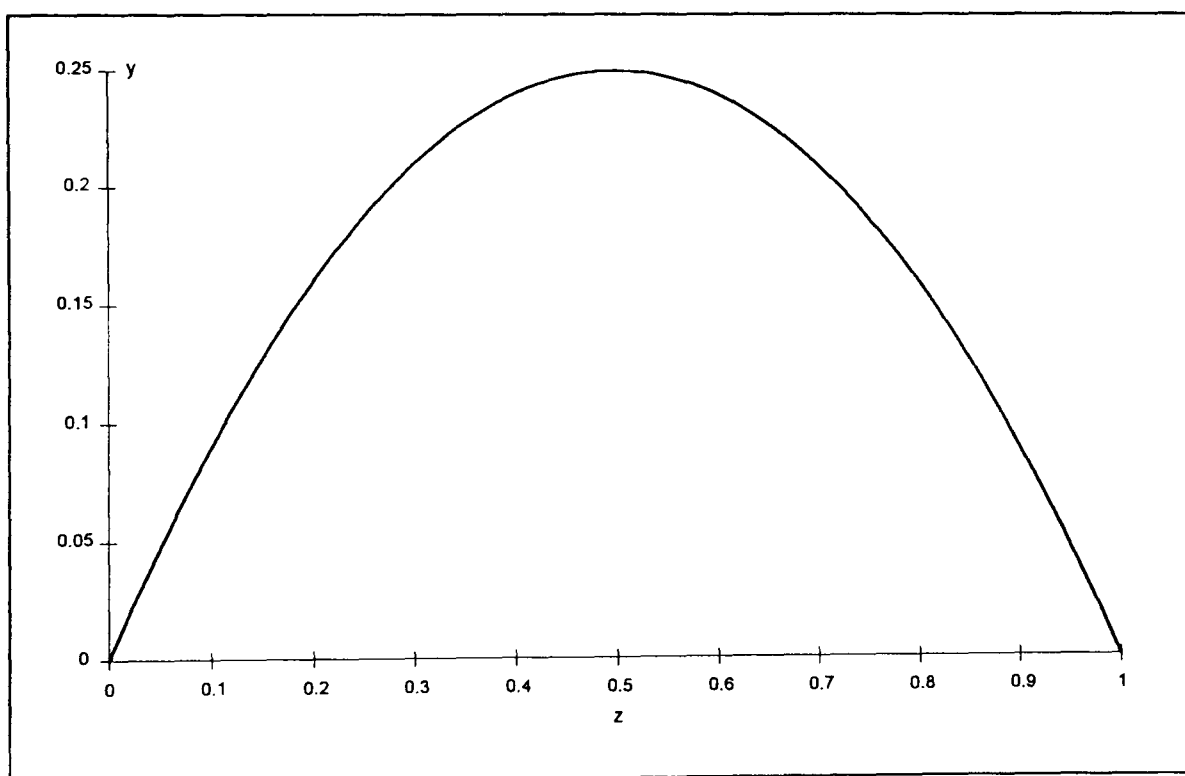


Figure 3.5: The contribution of the nodal output (z) to the error function.

This graph attains a maximum value of $y = 0.25$ when $z = 0.5$. This represents the value of z which produces the maximum modification to the weights. This is interesting because it will be recalled from figure 3.4 which described the sigmoid threshold function, that the output of a node is 0.5 when the sum of its inputs is 0. If the weights of the NN are well randomised before training commences then the sum of the inputs will be 0 and the

conditions for maximum weight adjustment are fulfilled. As training continues, the weights will be adjusted to produce nodal outputs tending to the desired values, 0 or 1. But to which extreme the output tends will depend on the training examples themselves. Therefore, the situation should never arise when the output, $z(j)$, is 1 but is desired to be 0 (similarly for $z(j) = 0$ and $d(j) = 1$). This condition could only conceivably be reached as a result of inconsistencies in the training set. This finding suggests that if the inconsistencies are few, they will make little difference to training because their measured error will be 0 and will lead to no weight adjustment. In this way, the NN will tend to the majority view. This finding also highlights the importance of the randomisation process of the NN weights. If large values are assigned then it becomes more likely that the sum of the inputs to a node will be, less than -3 or greater than 3 which would be enough to saturate the node to either 0 or 1 (figure 3.4). If this happened then training times could become much longer than necessary if a saturated node was desired to be the opposite polarity. To prevent this, all investigations began with the weights of the NN randomised to small, signed values ($-0.01 < w < 0.01$) to ensure that the sum at all nodes did not cause saturation and tended to the conditions for maximum weight adjustment at the start of training.

Selective data presentation strategy.

The training times for the backpropagation NN can become excessively long for large training sets. To overcome this, a data selection strategy, (Allred and Kelly, 1990) was incorporated into the backpropagation algorithm. This enhancement improves learning times by modifying the network weights only for those input examples which produce large output errors. No adjustment is made for those examples which produce outputs 'close' to the desired output. Specifically, a threshold variable λ was assigned, $\lambda = 0.75$

A training example was selected and processed through the NN. For each output, j , the absolute error, e_j , between the actual output and the desired output was calculated using equation 3.11.

$$e_j = |z(j) - d(j)| \quad (3.11)$$

An adjustment of the weights was performed when any output error was $> \lambda$.

When the error, e_j , for 90% of the training examples was less than λ , the threshold was reduced to 75% of its previous value and the process repeated. This cycle was continued until $\lambda \leq 0.1$ (when for 90% of the training set, every output required to be 1 was actually ≥ 0.9 , and every output required to be 0 was actually ≤ 0.1), whereupon the data selection strategy was abandoned and a weight adjustment was made for every training example.

3.2.2 The exclusive OR problem.

A 3 layer NN with any dimensions (limited to the available memory) was implemented in the C - programming language and was validated by considering the exclusive OR (XOR) problem proposed by Rumelhart et al, (1986). The XOR truth table for two binary inputs is shown in table 3.1,

Input A	Input B	Output
0	0	0
0	1	1
1	0	1
1	1	0

Table 3.1: The exclusive OR truth table.

The minimum NN configuration required to solve this problem is a NN with 2 inputs, 2 hidden nodes and 1 output (dimensions 2 x 2 x 1). A NN with these dimensions was trained to solve the XOR problem using the developed software. The training curve obtained is shown in figure 3.6, where the average error between the actual output and the desired output was plotted for each 1000 random presentations or epoch.

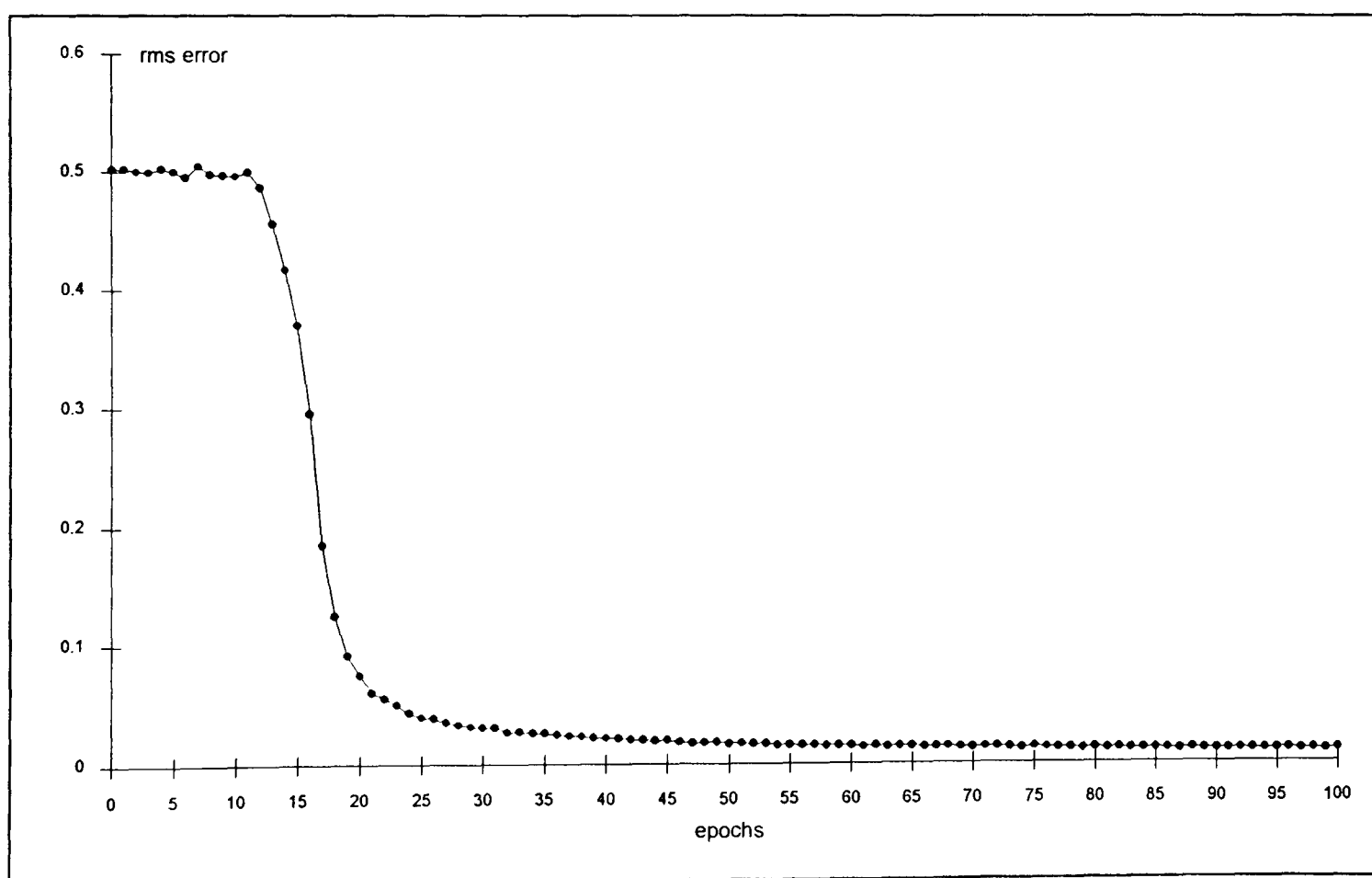


Figure 3.6: NN Learning Curve for solving the XOR Problem.

The averaged error at the output begins at 0.5, which confirms that the weights were suitably randomised. There was little change in the output error for the first 10,000

presentations (10 epochs) during which the weights are adjusted from their very small starting values to the larger values required for the solution. After this point, the output error falls rapidly as the NN converges to the final solution. From 25,000 presentations onwards, there is only a 5% error between the actual output and the desired output after which further convergence is small and slow. The state of the NN weights after 100 training epochs are shown in figure 3.7 and can be used to verify that this is a correct NN solution for the XOR problem.

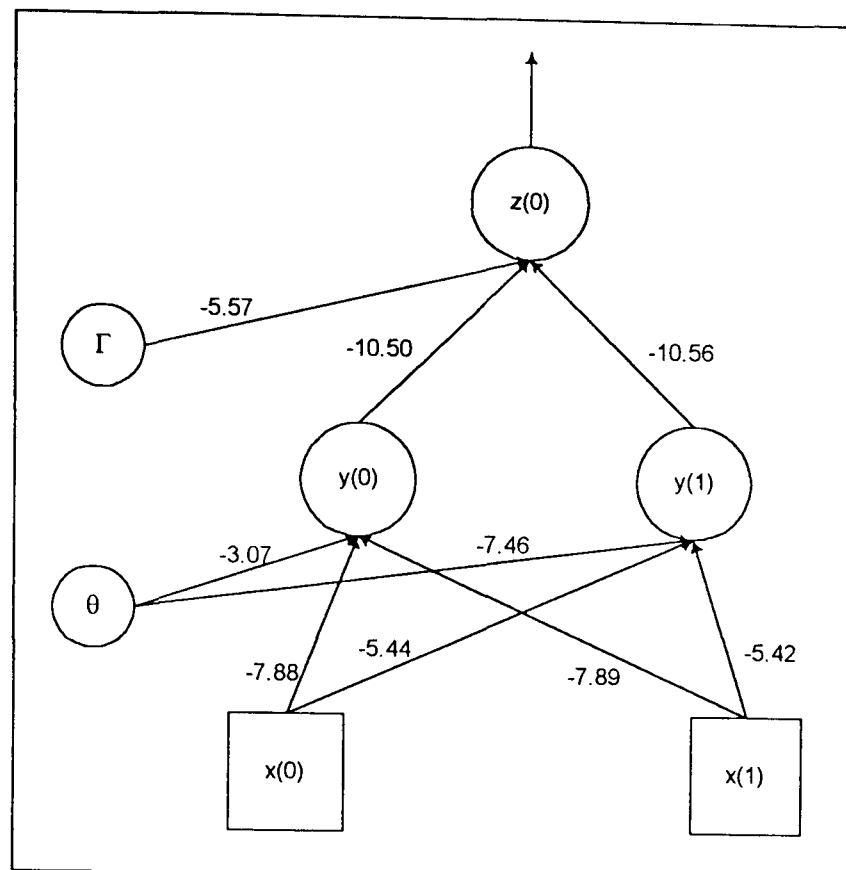


Figure 3.7: Neural network solution to XOR problem.

3.3 A neural network to classify the complete cardiocogram.

3.3.1 Method.

The first investigation involving NNs was to attempt to mimic the way CTGs are assessed by clinicians. It was proposed to train a single NN to fully classify all features from segments of CTG simultaneously. As certain features can last several minutes, the minimum clinically useful segment length was considered to be 5 minutes.

The contraction channel had a sampling frequency of 1 Hz, which gave 300 samples per 5

minute segment. The fetal heart rate varies so the number of samples per segment was not constant. If an average heart rate of 200 beats per minute were assumed (the highest likely), then 1000 samples per 5 minute segment would be obtained. Direct use of this data would require 1300 NN inputs, an impractical size. This data was therefore compressed by producing an averaged sample pair (1 heart rate sample and 1 contraction sample) for each 4.444 seconds of recording. This was a convenient figure as sample pairs were already averaged in intervals of 2.222 seconds for display purposes. Thus, for each 4.444 second interval a heart rate and contraction sample pair were produced, which gave 68 pairs or 136 data values for each 5 minute segment (figure 3.8).

A visual assessment of the averaged data showed that this averaging process altered the clinical interpretation of the high frequency heart rate variability which was therefore unsuitable for classification with this NN. It did not modify the clinical interpretation of any other feature. The expert classified the remaining three features of the CTG (baseline heart rate, accelerations in heart rate and decelerations in heart rate) into the sub-classes shown in figure 3.8 which required 11 output nodes. For each segment presented to the NN, one baseline output node was desired to be 1.0 (with the remaining nodes = 0), to indicate the range in which the baseline heart rate lies. A similar approach was used to classify decelerations. The presence of accelerations was indicated when the associated output node was 1.0. The number of middle nodes chosen was 20 to produce the largest size NN that could be simulated on our computer. The dimensions for the network were therefore 136 x 20 x 11 (figure 3.8).

The training set.

Compiling the training set is unquestionably the most important stage in the development of a NN solution. The training set must fully represent the entire problem, because otherwise the NNs performance may become unpredictable for patterns not included. This could result in the NN being unable to classify some patterns, or more seriously (bearing in mind the proposed application) erroneously classifying others. To address this problem, software was developed to allow an expert to extract 50,000 x 5 minute segments of CTG from a large database stored on optical disc media. However, it was considered that even a training set this large was unlikely to encompass all possible patterns.

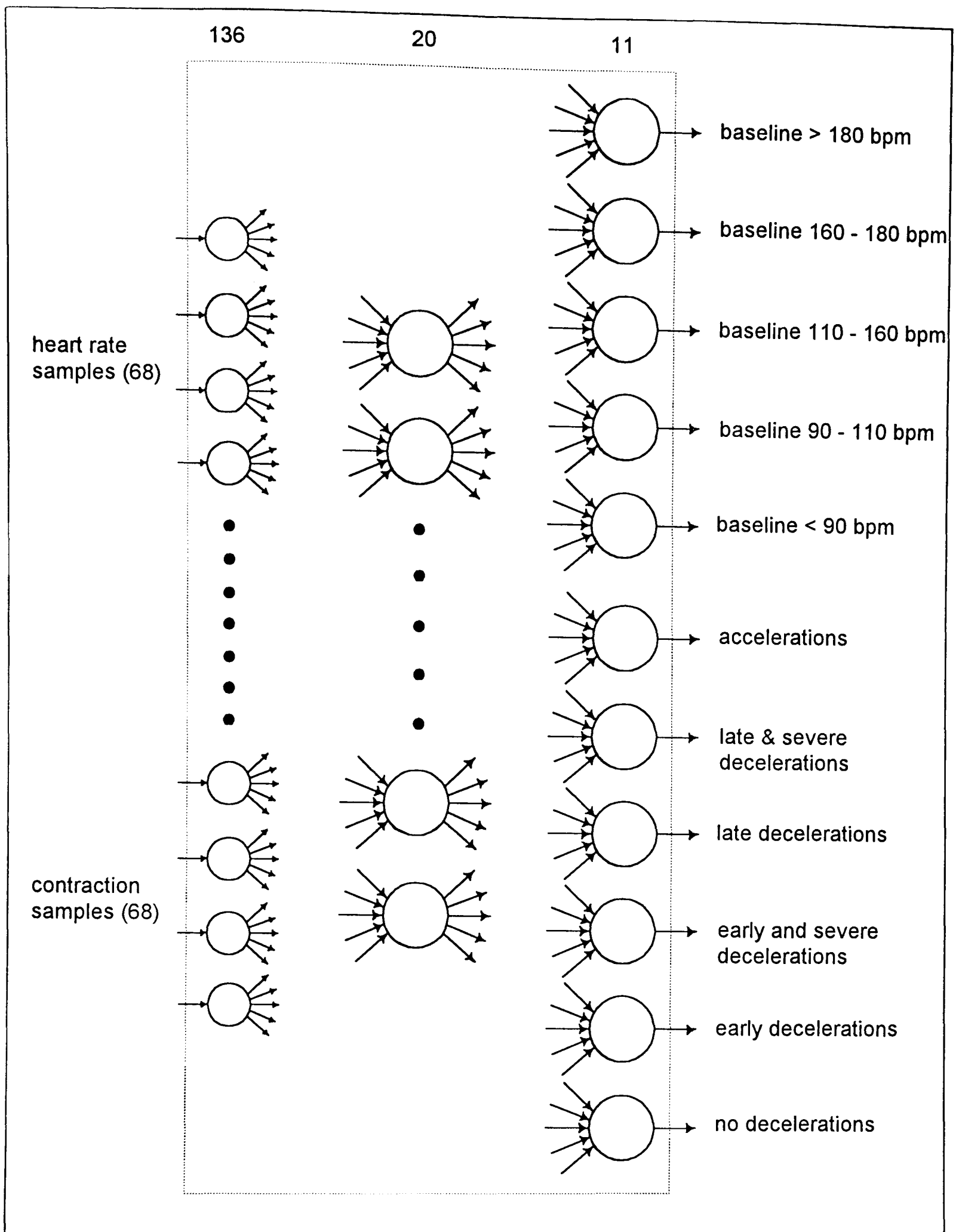


Figure 3.8: Neural network configuration for complete classification of the CTG.

3.3.2 Results.

The NN was trained on this data for 2 months continuously which represented over 100,000,000 data presentations. However, the network made little progress toward convergence even though the data selection algorithm was incorporated. Training was therefore terminated. This suggested that the problem in its current form was unsuitable for a NN solution and pre-processing was required to break the problem down into smaller, more manageable sub-tasks.

3.4 A neural network to classify the magnitude and timing of heart rate decelerations.

3.4.1 Method.

The problem was reduced by training a NN to classify just the magnitude and timing of heart rate decelerations in relation to uterine contractions, which have been shown to be the most difficult features to classify (Lindegard, 1992). The same input configuration as the previous approach was used together with 15 middle nodes and the 5 output deceleration classification nodes (136x15x5) as shown previously in figure 3.8. A training set of 1300 potential decelerations in heart rate were created by the expert, where a potential deceleration was defined as a transient departure below the baseline of more than 5 beats per minute for greater than 5 seconds. To add a degree of consistency to the task, the heart rate samples forming a deceleration were centralised to the NN heart rate inputs as shown in figure 3.9.

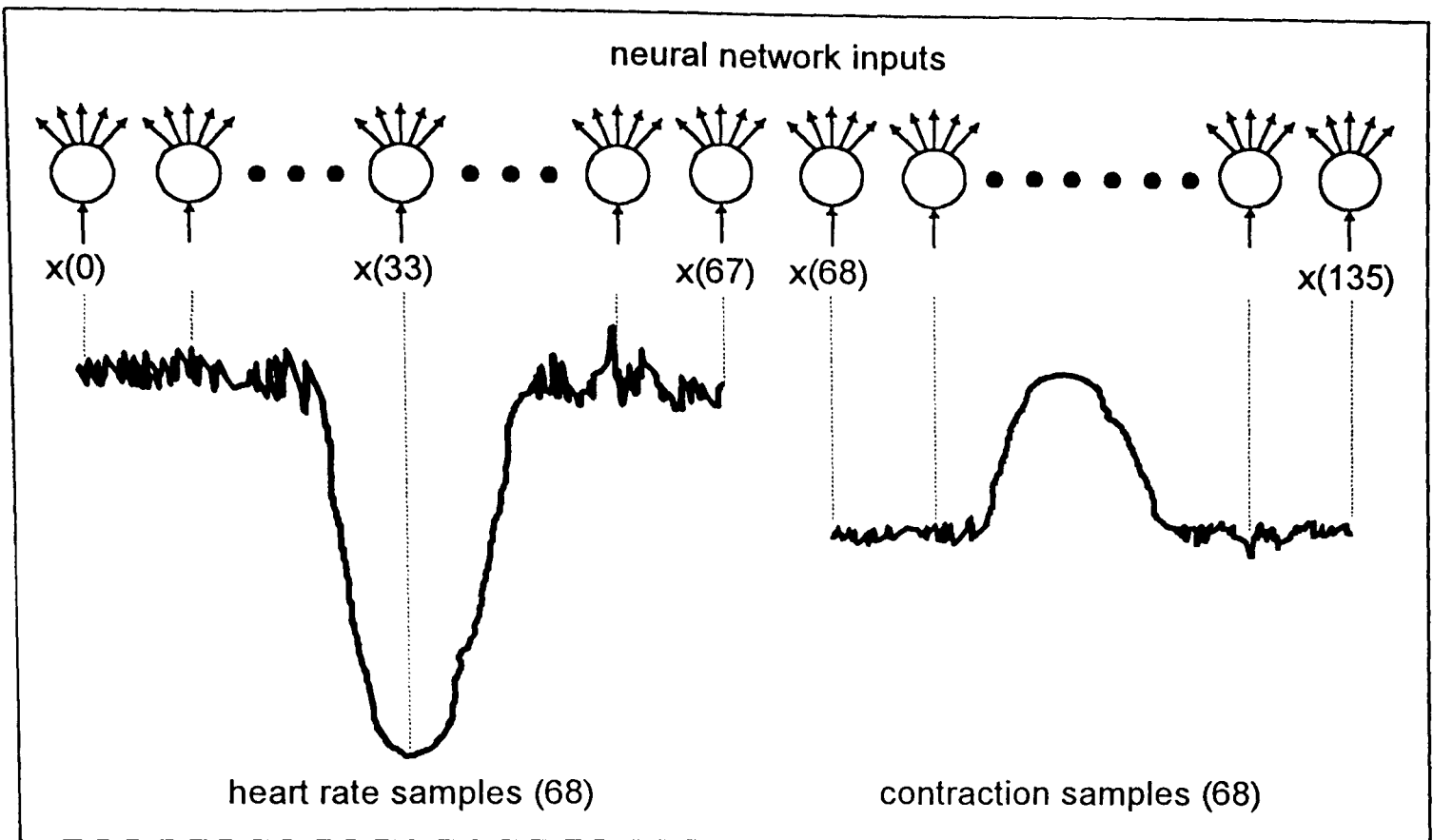


Figure 3.9: Centralisation of heart rate samples to neural network inputs.

3.4.2 Results.

The NN was trained until there was less than a 1% error between the expert classification and the NN classification for 90% of the examples. The training curve obtained is shown in figure 3.10. Convergence was achieved in approximately 24 hours.

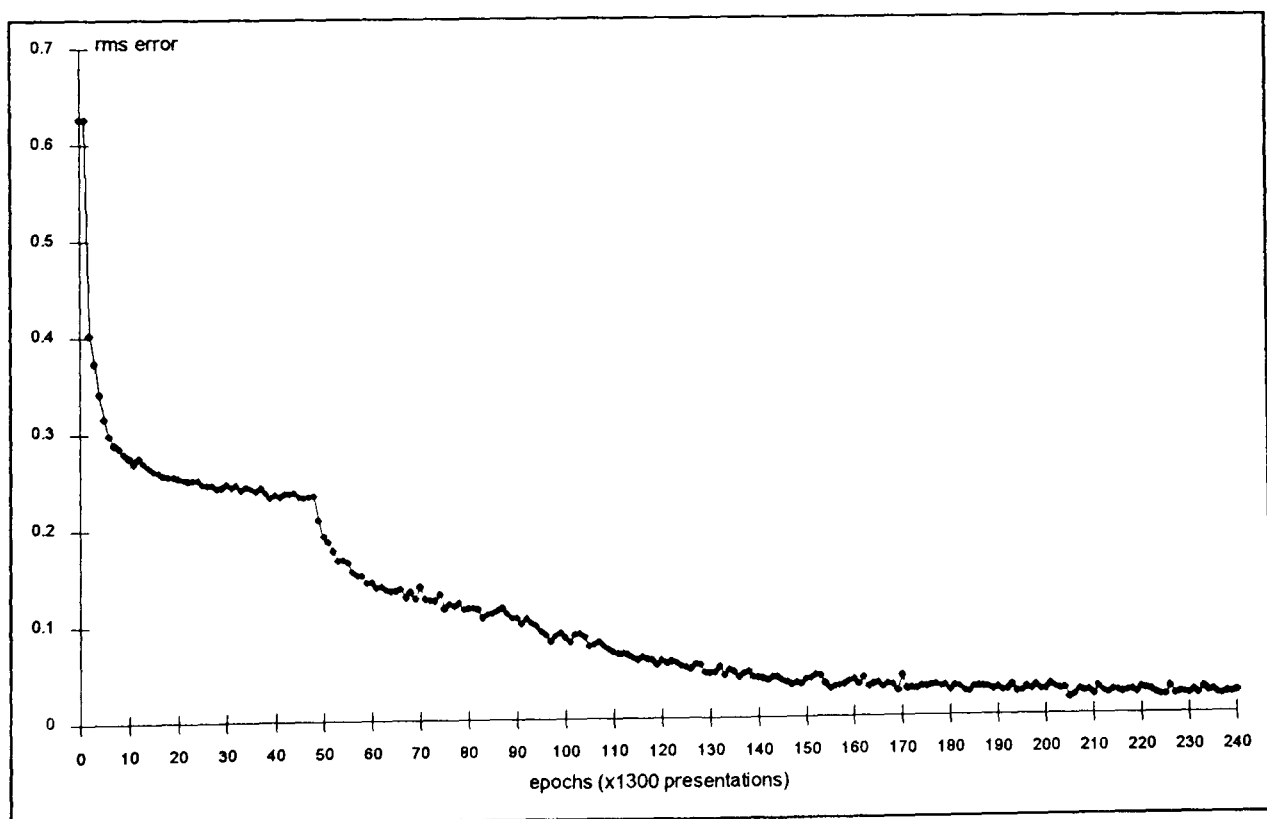


Figure 3.10: Training curve for classifying the magnitude and timing of decelerations.

The performance of the NN was assessed by an expert from a visual review of the NN's classification of decelerations within 56 previously 'unseen' labour recordings. The expert concluded that the NN performed well when the quality of the data recording was good. However, when the contraction data was of poor quality or absent, the NN performed poorly. This approach was considered of little practical use as clinically the signal quality during labour is inconsistent and often results in the contraction trace being poorly formed. This investigation indicated the difficulties of using actual data samples as input to the NN and suggested further pre-processing of the input data was required

3.5 A neural network to classify the magnitude of decelerations.

3.5.1 Method.

A NN using pre-processed input data was trained to classify the magnitude of decelerations only. In practice, the timing of a deceleration in relation to the contractions would have to be subsequently classified by an algorithm. The selective data presentation algorithm was unnecessary because the training times were short.

Previous computerised approaches to identify decelerations (Krause, 1990; Maeda, 1990; Dawes et al, 1991) have largely been based on the relationship between the depth below the baseline heart rate and the duration of the feature, figure 3.11.

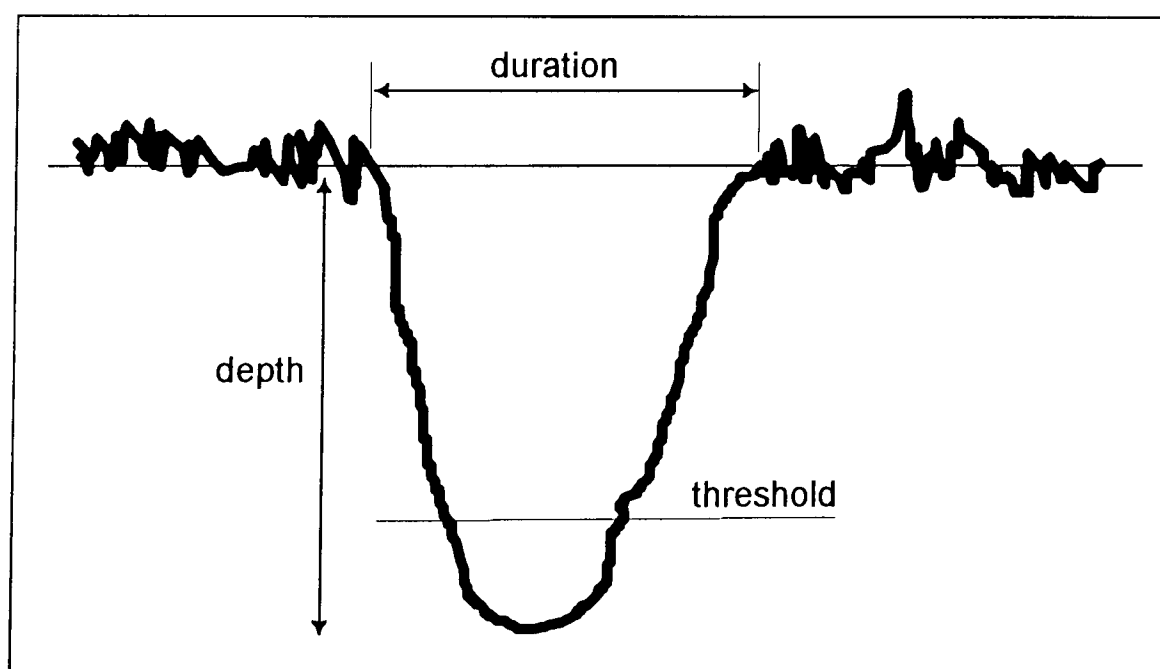


Figure 3.11: Conventional model for classifying the magnitude of decelerations.

The System 8000, (Oxford Medical Ltd, Abingdon, Oxford, UK) is the only commercially available computerised antenatal monitor which performs automatic feature extraction from the antepartum CTG. During labour, the CTG is a far more complex and is more difficult to interpret. The antepartum CTG is recorded over an interval of 20-30 minutes when the mother is relaxed and the fetus is not stressed. Put simply, if any abnormal features appear then there is cause for concern. Although the interpretation of the 2 waveforms are different, some of the features are similar, for example decelerations. The System 8000 identifies *decelerations* as those transient departures which fall below the baseline for more than 20 bpm for more than 30 seconds, or by more than 10 bpm for more than 1 minute. An identified *deceleration* is then further classified as '*large*' (or '*severe*') if the area is greater than 20 lost beats (Dawes et al, 1991). The suitability of this algorithm to make the initial classification of a deceleration was investigated by comparing the classifications of 1,000 examples by the System 8000 algorithm and two experts independently. The experts made the same classification for 86.1% of the features. The algorithm however, made the same classification as the experts for only 60.4% and 55.8% of the features respectively. It was likely that the relatively poor agreement between the System 8000 algorithm and the experts could in part, reflect the need for a more sophisticated classification model for features occurring during labour. An extension of the classification model was therefore proposed and is shown in figure 3.12.

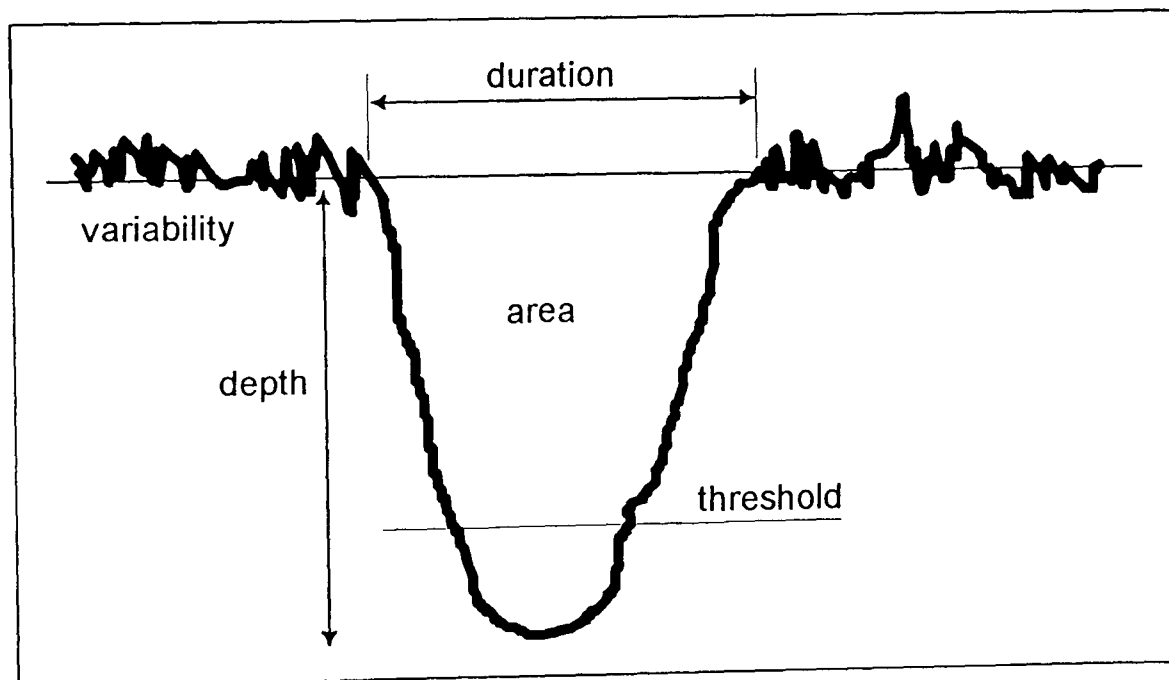


Figure 3.12: Extended model for classifying the magnitude of decelerations.

The area of a deceleration is used by the System 8000 algorithm to further classify *decelerations* as *severe decelerations*. It was the opinion of experienced clinicians that this shape information may also be important for making the initial classification. The heart rate variability was also thought to be important, as it was considered that during episodes of high variability the thresholds for classifying decelerations may need to be increased and during episodes of decreased variability, they may need to be reduced.

An algorithm was developed to identify possible decelerations and measure the four variables, area, depth, duration and variability which were used as the inputs to a NN. The NN required 3 outputs to classify the feature as either, *not a deceleration*, *a deceleration*, or a *severe deceleration*. A training set of 456 examples, evenly distributed across the three classifications, was compiled and classified by an expert. Selection was based on the requirement to broadly cover the different sizes and shapes of decelerations.

3.5.2 Optimising the number of middle layer nodes.

Defining the optimal number of middle nodes to incorporate in a NN is not obvious. Therefore 10 NNs with dimensions 4 x n x 3, where n = 1 to 10, were trained separately on this data. Each network was allowed to train for 1,000,000 random presentations of the examples, which represented approximately 2193 presentations of each example. The percentage number of examples classified correctly was recorded for every 456 random presentations, or epoch. A correct classification was achieved when the output required to be 1 was actually greater than 0.9, and the outputs required to be 0 were less than 0.1. When training was halted, the mean and variance of the percentage number of correctly classified training examples from the last 10 epochs were calculated for each of the 10 NNs and are shown in table 3.2.

Number of middle nodes (n)	mean (%)	variance
1	29.6	6.55
2	78.3	15.34
3	78.1	21.45
4	79.6	16.45
5	82.1	2.37
6	84.3	4.74
7	83.5	3.26
8	83.4	2.25
9	83.7	3.77
10	82.9	5.16

Table 3.2: Mean and variance of correct classifications made during the last 10 epochs of training for NNs with 1 to 10 middle nodes.

NN₁ (where subscript 1 denotes the number of middle nodes), achieved only 29.6% correct classifications for the training set and so was unsuitable. The remaining networks fell into two groups; NN₂ to NN₄ have high variances compared to NN₅ to NN₁₀ which implies they had formed a less stable solution. A one way analysis of variance showed that there were no significant ($\alpha = 0.05$) differences in the mean number of examples classified correctly by NN₅ to NN₁₀. Hence NN₅ was identified as optimal, as it achieved

convergence with the fewest middle nodes and was therefore the most computationally efficient.

3.5.3 Test of generalisation.

The classification of decelerations is subjective and so there are no 'gold standards' to assess the performance of NN₅ to correctly classify 'unseen' examples. For this reason, the performance of NN₅ and the System 8000 algorithm were compared with human interpretation. Six reviewers were recruited; two experienced senior registrars who interpret the CTG daily who were considered experts, an experienced research midwife, a midwifery Sister, a locum Consultant obstetrician/gynaecologist with previous labour ward experience and an engineer who had no experience in CTG classification.

Each reviewer independently viewed on a computer screen 232, 15 minute sections of heart rate recording, in which a possible deceleration was highlighted. However, unknown to the reviewers, these 232 sections of recording actually contained 116 unique examples which were presented twice. The 116 different examples were presented consecutively and then re-presented in a randomised sequence.

Each participant was asked to classify the magnitude of the 232 decelerations as, *not a deceleration*, *a deceleration*, or a *severe deceleration*. These examples were also presented to the NN₅ and the System 8000 algorithm. The NN₅ was deemed to have successfully classified a feature when 1 output only was greater than 0.8 and two outputs were less than 0.2.

3.5.4 Results.

NN₅ classified all 232 decelerations. Table 3.3 shows the level of agreement (as a percentage) between the experts, the System 8000 and the NN₅ as well as each reviewer's consistency. For example, reviewer C consistently classified 79.3% of the examples and agreed with 73.7% of reviewer B's classifications.

	A	B	C	D	E	F	8000	NN ₅
A	89.7	81.0	74.5	65.1	56.9	62.9	56.5	75.0
B	81.0	97.0	73.7	72.9	58.6	58.2	68.9	81.9
C	74.5	73.7	79.3	60.8	56.5	60.8	56.0	75.4
D	65.1	72.9	60.8	84.5	59.1	55.2	68.1	58.6
E	56.9	58.6	56.5	59.1	76.7	49.1	52.6	54.3
F	62.9	58.2	60.8	55.2	49.1	65.5	38.8	54.7
8000	56.5	68.9	56.0	68.1	52.6	38.8	100	46.6
NN ₅	75.0	81.9	75.4	58.6	54.3	54.7	46.6	100

Table 3.3: The reviewers agreement.

Reviewers A, B and C, the senior registrars and research midwife respectively, all had a high level of internal and external consistency which demonstrates their expertise for this task. Reviewer D the engineer, had a high level of internal consistency but with rather lower inter-agreement. The System 8000 algorithm did not agree particularly well with any reviewer. NN₅ however, shows a high level of agreement with A, B and C, and little agreement with D, E and F or with the System 8000 algorithm.

The poor performance of the System 8000 algorithm was investigated. Table 3.4 shows the breakdown in classifications made by observer B, the most consistent expert and the algorithm.

	System 8000		
Reviewer B	Not Deceleration	Deceleration	Severe Deceleration
Not Deceleration	92.9	2.8	4.2
Deceleration	54.5	36.6	8.9
Severe Deceleration	0.0	6.0	94.0

Table 3.4: Breakdown of classifications between expert B and System 8000 algorithm.

The algorithm agreed well with reviewer B for the classification of *not decelerations* and *severe decelerations*. However, the algorithm only agreed with 36.6% of the features classified by B as *decelerations*; 54.5% of the features classified as *decelerations* by B were classified as *not decelerations* by the algorithm. These results were consistent for A and C. This insensitivity may result from the initial requirement that a feature is a *deceleration* only if the heart rate falls below the baseline by more than 20 bpm for more than 30 seconds, or by more than 10 bpm for more than one minute. It was found that the performance of algorithm could be improved if the initial classification was taken as a fall in fetal heart rate

of more than 10 bpm below the baseline for more than 15 seconds with an area greater than 5 lost beats. The subsequent classification for *severe decelerations* was unchanged. Comparing this modified algorithm with the other reviewers obtains the results in table 3.5.

	Reviewer							
	A	B	C	D	E	F	8000	NN ₅
Modified System 8000 algorithm	72.1	76.7	70.0	73.3	52.6	48.3	63.4	72.4

Table 3.5: Agreement table with modified System 8000 algorithm.

This modified System 8000 algorithm obtained a more general solution to the problem. It obtained good agreement with, A, B and C, the NN₅ and interestingly, with D (the engineer). The NN₅ agreed with reviewers A, B and C but especially with B, the reviewer who classified the training set. It may be that the algorithm could be further improved if the heart rate variability preceding the feature were also taken into account.

3.6 Discussion of the suitability of neural networks for feature extraction.

Three investigations have examined different ways in which NNs might be implemented to extract features from the CTG during labour. The first investigation was unsuccessful in training a NN to simultaneously extract and classify all features of interest from a segment of averaged raw data samples. This was perhaps because of the numerous patterns possible and the lack of consistency in the relative positioning of these features within each segment. The second approach reduced these problems by training a NN to classify a single feature, again using averaged raw data samples as input. A numerical algorithm was used to identify the possible features in the waveform which were then presented to the NN with the same relative input positioning to ensure a degree of consistency. This approach produced a solution for the training samples but when tested on 'unseen' data, it was found to be unable to classify features during periods of poor signal quality.

The most effective implementation was obtained when a numerical algorithm identified a possible feature and pre-processed the data to quantify the important parameters which then formed the input to the NN. This method was able to obtain a solution with a performance comparable with experts. In addition, this approach had the advantage of reducing the NN dimensions and ensured that only meaningful information was presented.

It is important that some care is exercised when using NNs in sensitive applications such as

clinical decision support. The examples chosen to train a NN will largely determine its eventual performance. If the training set does not contain a comprehensive representation of the problem domain, then the NN will either be unable to classify certain future examples or more seriously, perhaps even classify some examples wrongly. Once trained, the network must be tested on a second set encompassing the full range of possible examples to ensure its safe future operation.

The application of a single NN on its own, for high level tasks such as suggesting clinical actions, may also be unsuitable. This lies in the fact that they are conceptually 'black boxes' which produce an output in response to a mathematical manipulation of a given input. It is this property which makes NNs desirable because it removes the need to formulate expert knowledge. However, this process also precludes an explanation of the reasoning which led to a certain course of action being recommended.

This study suggested a hybrid approach would be the most suitable implementation for the intelligent fetal monitoring system. Numerical algorithms could be used to extract features and subsequent algorithms or small NNs could then be used for classification.

3.7 Baseline heart rate estimation.

There are two features associated with the baseline fetal heart rate. The first is the dynamic baseline reference waveform which can be drawn through the data and about which other features are classified. The second, which can be obtained from the reference waveform, is a single value which represents the general proximity of the waveform. Consider figure 3.13 for example.

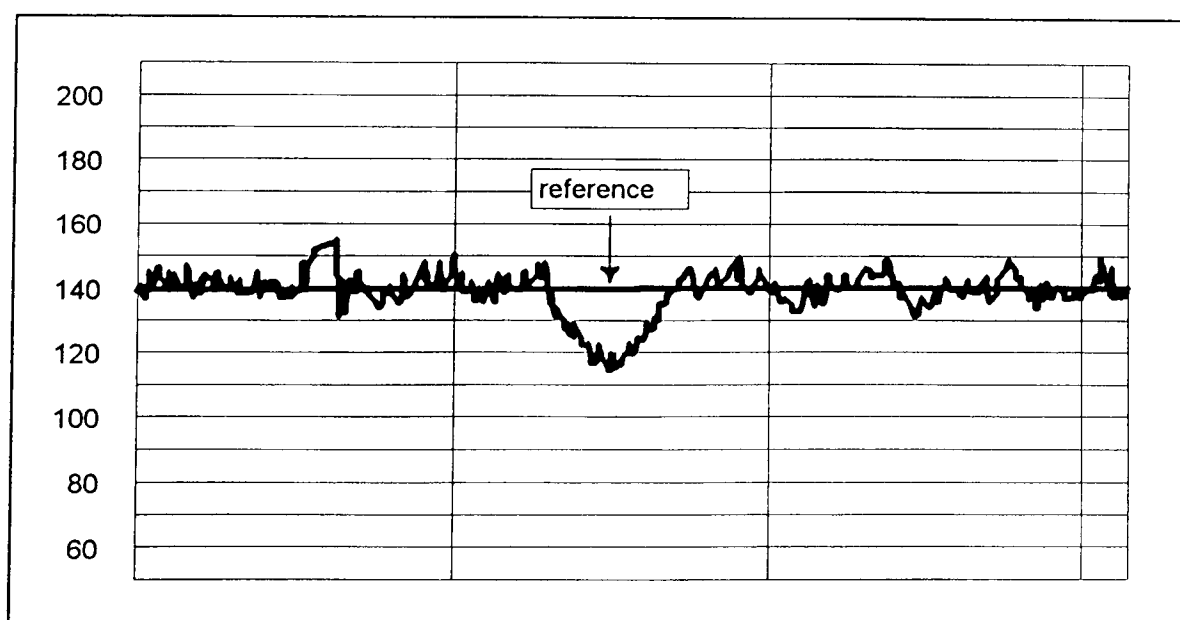


Figure 3.13: Baseline heart rate estimation.

Figure 3.13 shows the imaginary reference waveform drawn through the heart rate pattern which has a baseline heart rate of 140 bpm. This reference line is fundamental to the correct interpretation of the CTG because the classification of other features depend upon it. It will also be noted from figure 3.13, that the baseline is unaffected by the large deceleration in the centre of the recording. This is one of its most important properties.

There are certain patterns where it can be difficult to establish the precise location of the baseline heart rate. Consider figure 3.14 for example. Does this pattern have a baseline > 160 bpm (baseline 1) or is it approximately 110 bpm (baseline 2)?

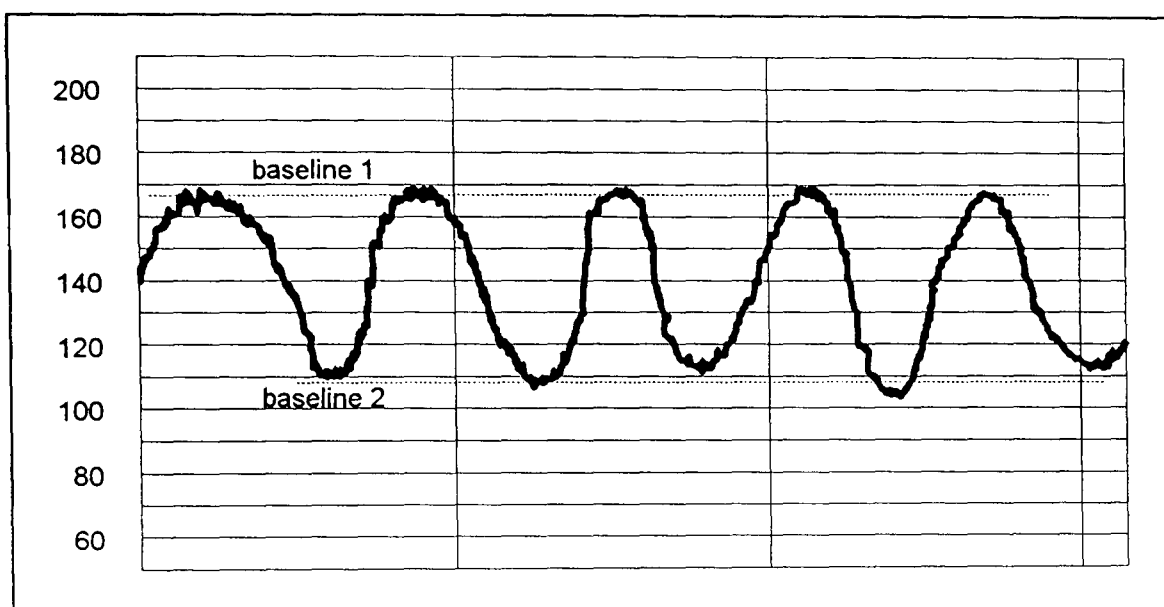


Figure 3.14: Unstable baseline heart rate.

If greater than 160 bpm, then this CTG contains decelerations. On the other hand if the baseline were taken as 110 bpm, then the features would be classified as accelerations. The difficulty is that the subsequent management of the case may depend on which interpretation is accepted. In reality, this pattern would be recognised as abnormal by virtue of the fact that the baseline is so unstable and it may be argued that this pattern does not have a baseline. But this example demonstrates that the algorithm required to obtain the baseline heart rate must be accurate and robust, otherwise all subsequent classifications and interpretation could be misleading.

It became clear from consultation with experts that when they assess baseline, they tend to look for the periods when the heart rate is most stable and join these together with an imaginary line. The stable periods are those to which the eye is most naturally drawn. In figure 3.13 it was simple to draw this line because despite some short lived deviations, the heart rate was most stable at 140 bpm. In figure 3.14, the heart rate pattern was seldom stable but the periods of greatest stability occurred at the peaks and troughs. As there were two regions of relative stability, it was difficult to decide which to accept. Periods of stability are indicated by the average heart rate about which the given recording

fluctuates. However, this is not the mean of samples because this would be influenced greatly by large deviations. The fact that the baseline is not influenced by transient departures above or below it, suggested that the best approximation would be given by the most 'fashionable' value, or the modal average. An algorithm was developed to calculate the baseline heart rate using this approach and took the following form.

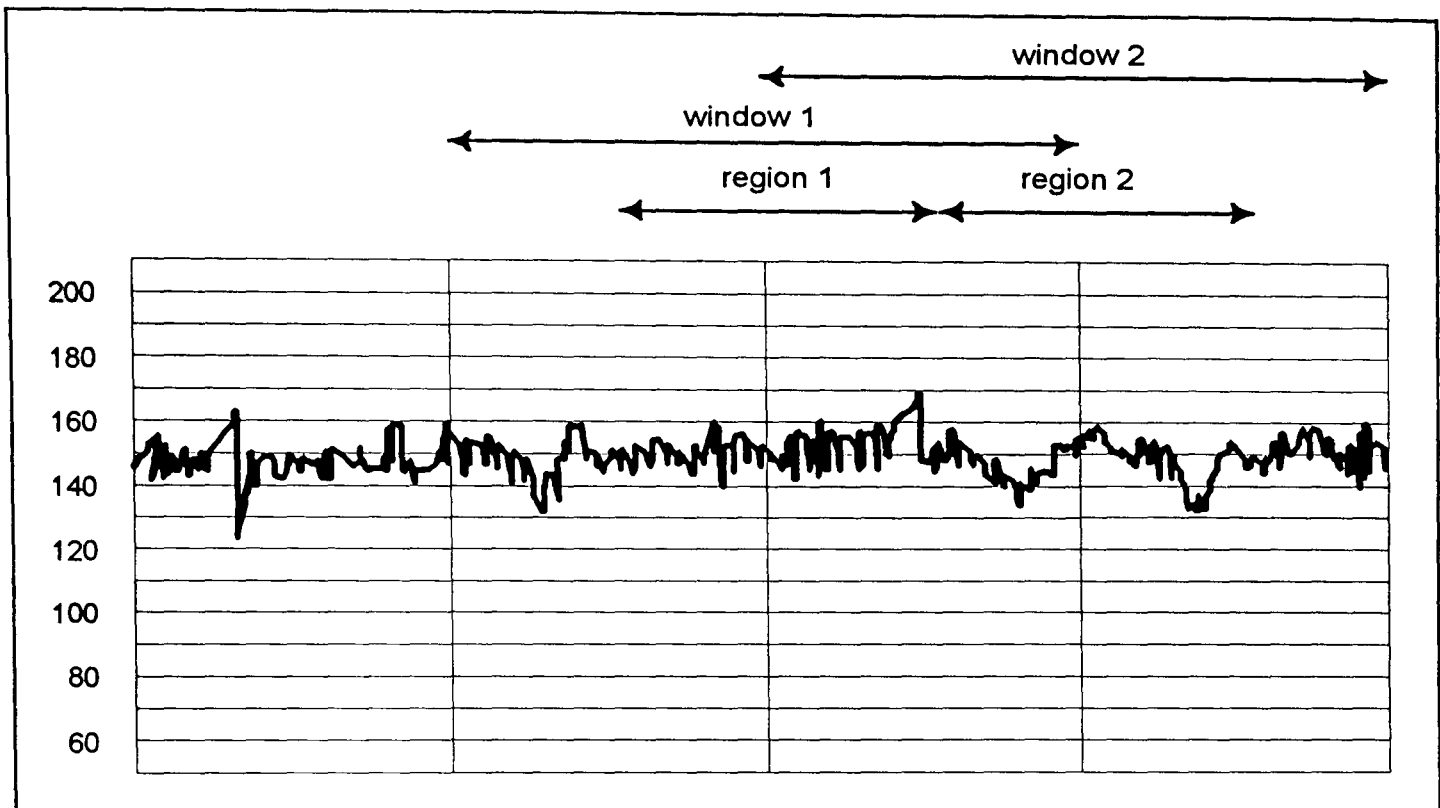


Figure 3.15: Baseline heart rate estimation.

With reference to figure 3.15, consider the baseline for region 1, a 5 minute period of recording.

1. The statistical frequency distribution of the heart rate data samples contained in the 10 minute window, window 1, was formed (figure 3.16). This window overlapped region 1 by 2.5 minutes at each end. This improved the algorithm because it allowed the baseline value for region 1 to be influenced by past and 'future' samples.

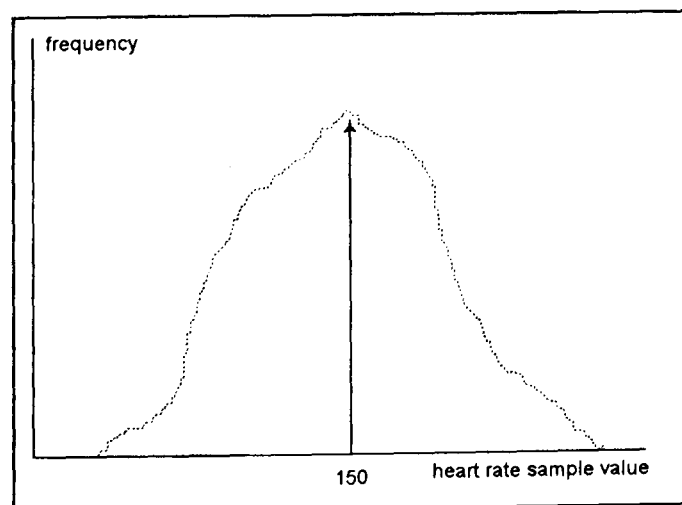


Figure 3.16: Statistical frequency distribution of heart rate values.

The number of samples, x_j , with a heart rate value, hr , is given by,

$$\sum_{i=1}^N x_i(hr)$$

Where N is the total number of samples in the window and, $i = 1, 2, \dots, N$.

2. The frequency distribution considered heart rates between 1 and 210 bpm (although heart rates of less than 50 bpm were not expected). The most densely populated range of heart rate values was identified; the sum of the number of heart rate values in the ranges $1 \leq hr < 10$, $2 \leq hr < 11$, ..., $201 \leq hr < 210$, were found. The range which contained the greatest number of samples was taken as the range in which the baseline was most likely to be found. This process is described by equation 3.12.

$$modal\ range_{hr \rightarrow hr+9} = \max_{hr} \left[\sum_{i=1}^N x_i(hr) \right] \quad (3.12)$$

Where, $hr = 1, 2, \dots, 201$.

3. The modal heart rate of this modal range gives the baseline value for region 1. The process would then be repeated for region 2.

This procedure was adopted in preference to simply taking the mode of the entire frequency distribution for two reasons.

1. Noise in the form of short-lived signal drop-out maintains a constant non-zero value which is likely to be the modal value.
2. The heart rate pattern fluctuates about the baseline, but a heart rate value need not equal the baseline value. Therefore, the total number of heart rates equal to the baseline could be few and not necessarily the modal value.

The baseline can change from region to region. If this change is large, it could indicate fetal compromise. Under these circumstances, the algorithm would be required to respond quickly.

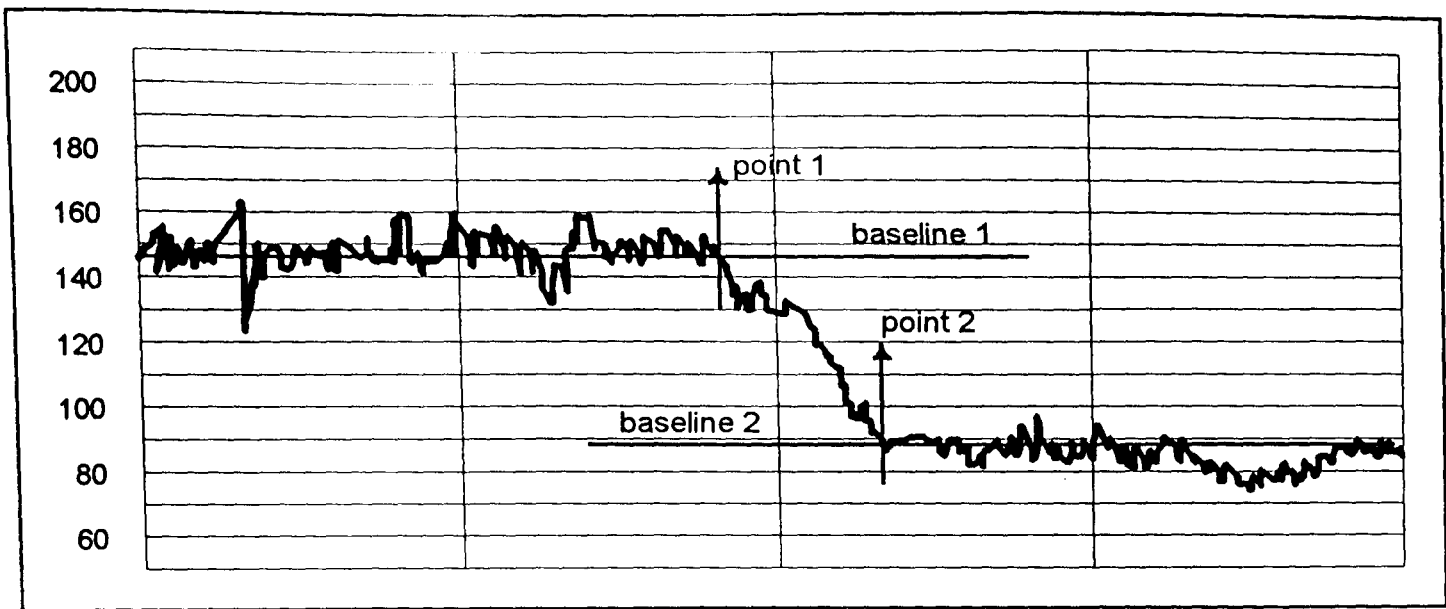


Figure 3.17: A baseline change.

In figure 3.17, the heart rate pattern falls from baseline 1 (> 140 bpm) to baseline 2 (< 90 bpm). Here, the algorithm was designed to identify point 1, where the heart rate diverges from baseline 1 and, point 2, where the heart rate converges to baseline 2. With these points identified, the equation of the line joining them was found which was taken as the transitional baseline heart rate.

3.8 Heart rate accelerations.

An acceleration is a relatively long-lived transitory increase in fetal heart rate from the baseline. The complete event can last up to several minutes but is often less. Several definitions for classifying accelerations have been suggested (Hon, 1968; Krebs et al, 1979; Dawes et al, 1982), which all rely on the model shown in figure 3.18. This considers the relationship between the duration of the feature with its maximum departure above the baseline.

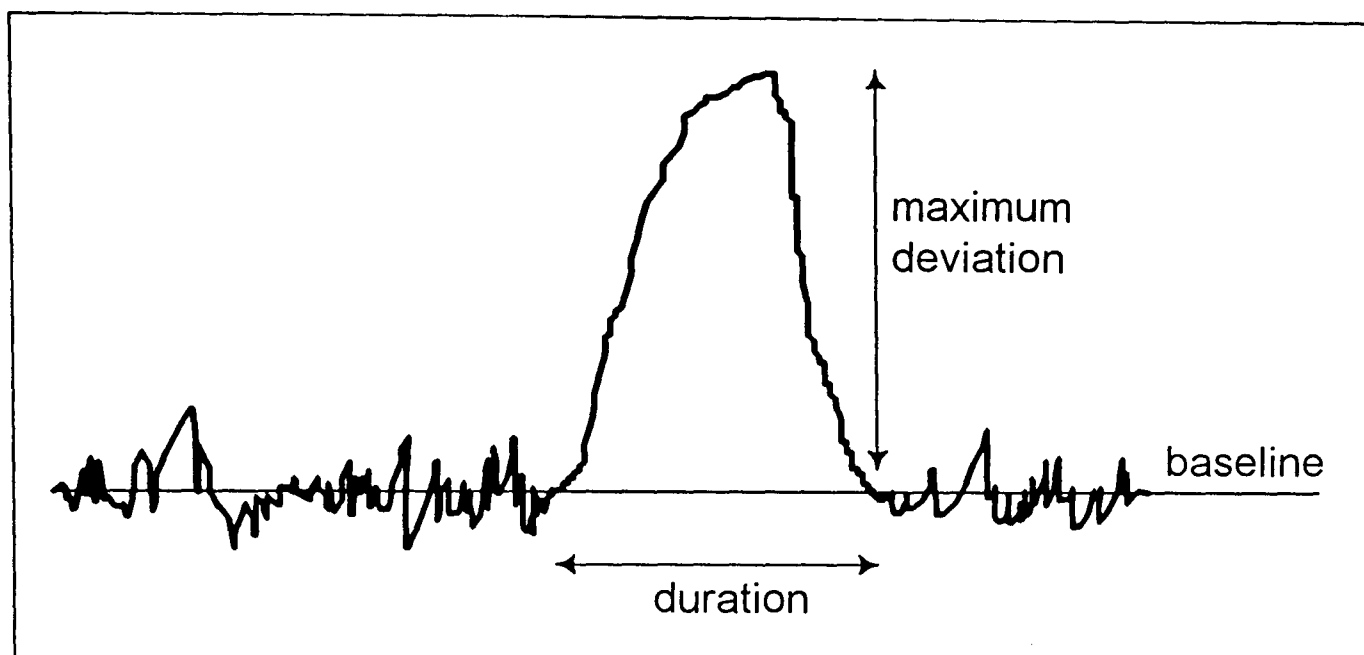


Figure 3.18: Classification model for accelerations.

The suitability of this model was investigated to establish whether it could be used to classify accelerations similarly to experts. 500 transient departures of at least 3 bpm above the baseline with a duration of at least 5 seconds, were classified as either, *not acceleration* or *acceleration*, by 2 experts independently. The experts agreed with each other for 88.3% of their classifications. The duration and maximum deviation above the baseline were plotted for the classified features. The features classified by one expert as *not accelerations* are shown in figures 3.19 and those classified as *accelerations* are shown in figure 3.20.

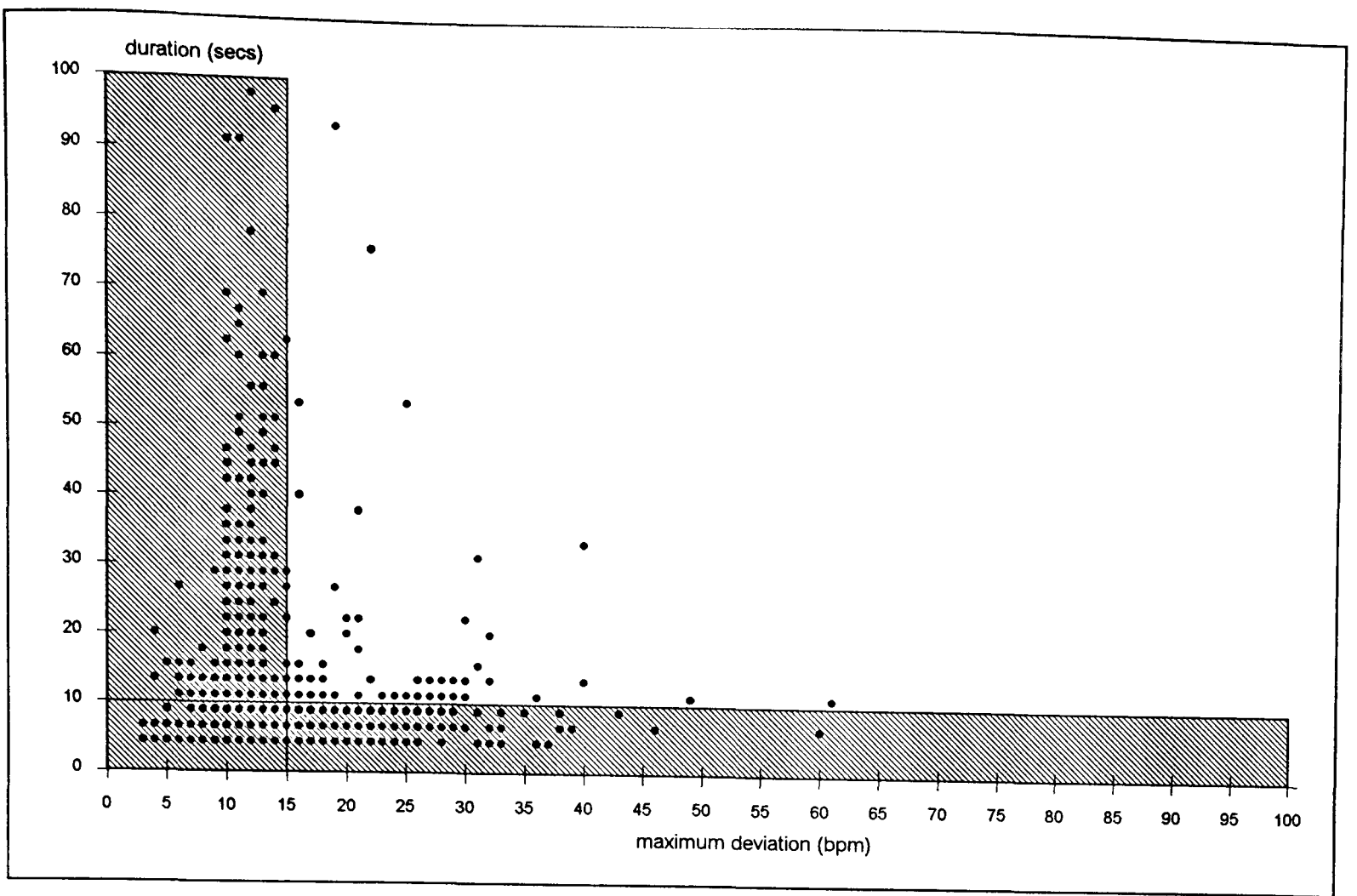


Figure 3.19: Features classified as *not accelerations* by 1 expert.

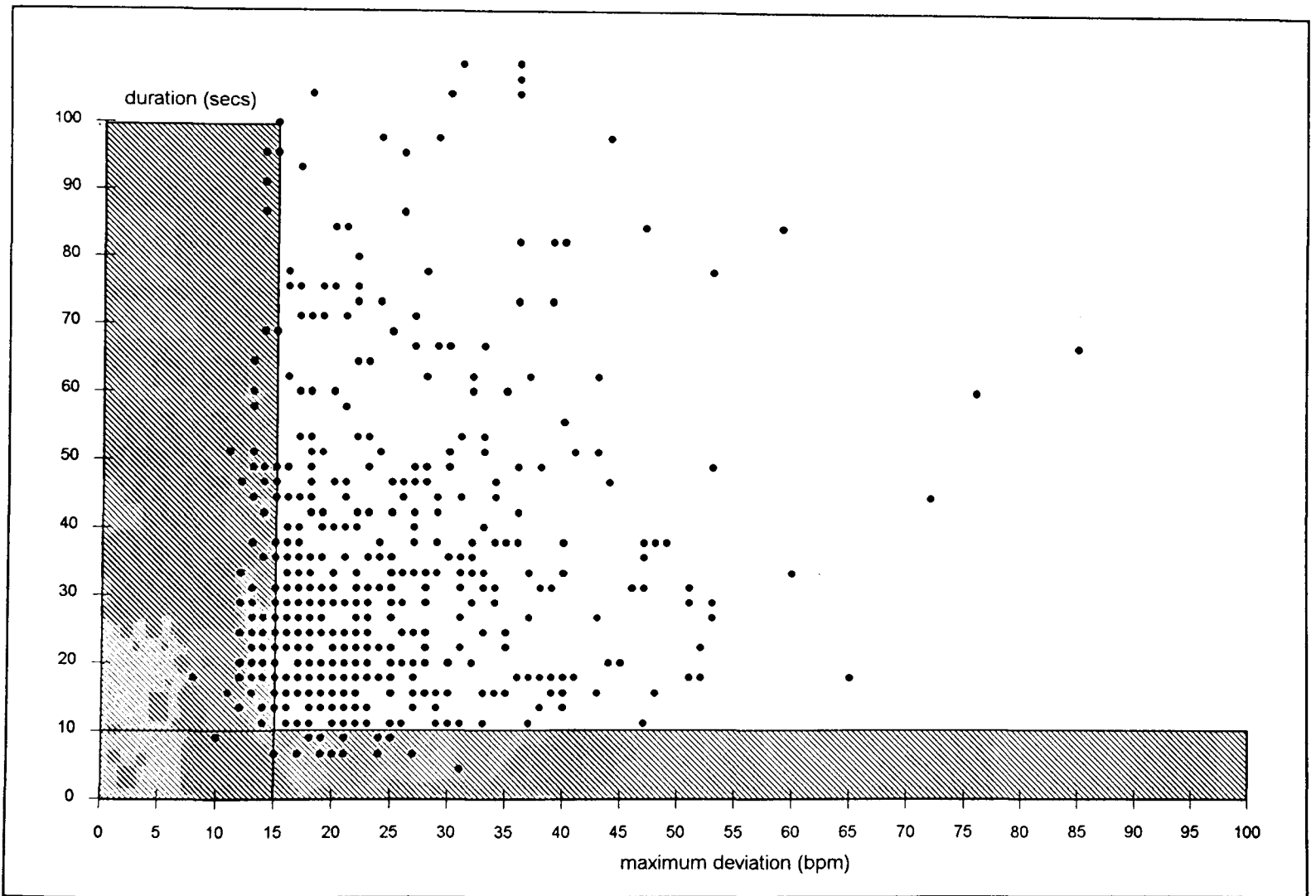


Figure 3.20: Features classified as *accelerations* by expert 1.

It was found that an agreement of 82.2% could be obtained with this expert if the threshold used in the classification model for the maximum deviation above the baseline was 15 bpm and the threshold for the duration was 10 seconds. The rejection region using these thresholds is indicated by the shaded area of the graphs. The features classified as *not accelerations* by expert 1 in figure 3.19 which fall outside the shaded region, are those that would be incorrectly classified using the proposed thresholds. Similarly in figure 3.20, the features classified as *accelerations* which fall inside the shaded region are those which would be incorrectly classified. These graphs do not give the most accurate representation of the density of classified features as those with the same deviation and duration are plotted on top of one another. However, the average agreement with both experts using this criteria was found to be 80.1% which was considered acceptable.

3.9 Heart rate variability.

Fetal heart rate variability is taken as the amplitude of irregular fluctuations about a baseline during stable recording between accelerations and decelerations. The baseline here, is not necessarily the baseline described in 3.7. It could be a local baseline specific to the region of stable recording being considered. The collective classification of all stable regions within a given segment of recording would then provide the overall heart rate variability classification.

The developed algorithm considered the CTG in 15 minute segments. All accelerations and decelerations were first identified and excluded from the analysis. A one minute assessment window was passed over the remaining stable data. For each complete one minute of recording, the data within the window was considered to perturb about a local baseline given by the straight line equation, $y = \beta x + \alpha$, where y_i is the heart rate associated with the $x_{i\text{th}}$ data sample. This method is represented in figure 3.21.

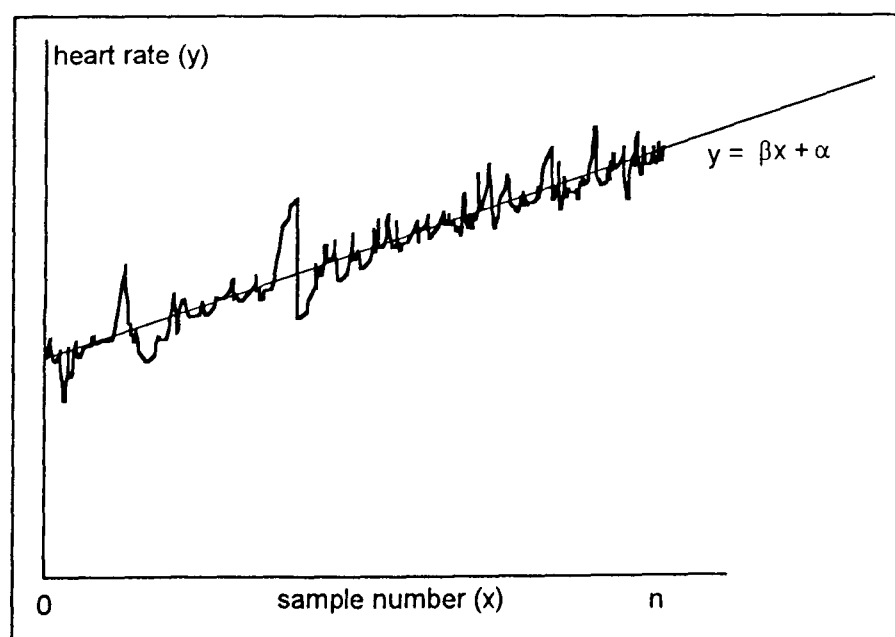


Figure 3.21: Heart rate variability model.

The equation of this local baseline was found using regression analysis (equation 3.13),

$$\beta = \frac{\sum_{i=0}^{n-1} (x_i y_i) - \frac{1}{n} \sum_{i=0}^{n-1} x_i \sum_{i=0}^{n-1} y_i}{\sum_{i=0}^{n-1} (x_i)^2 - \frac{1}{n} (\sum_{i=0}^{n-1} x_i)^2} \quad (3.13)$$

and $\alpha = \bar{y} - \beta \bar{x}$

where n is the number of data samples in the one minute epoch.

The equation of the local baseline was then subtracted from the heart rate samples to transform the data about a zero baseline (figure 3.22).

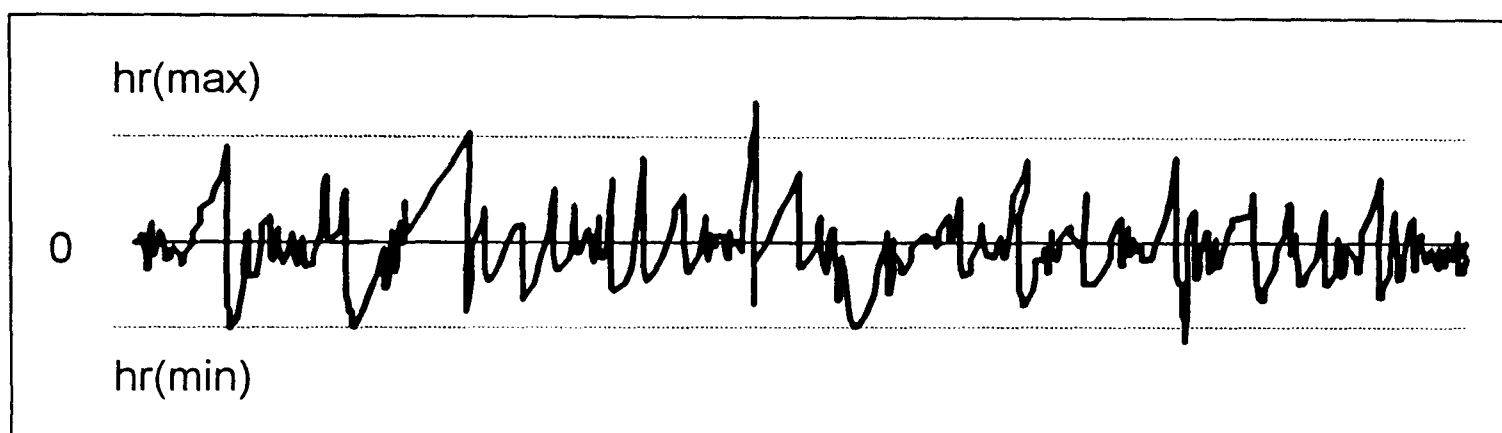


Figure 3.22: Heart rate data with local baseline removed.

The frequency distribution of these transformed samples was then constructed and the heart rate at the 5th percentile was taken as hr(min) and heart rate at the 95% was taken as hr(max). These were used in preference to the absolute maximum and minimum values to reduce the risk of a spurious data samples giving a false representation of the deviation. The difference between these heart rates gave the heart rate variability, hrv (equation 3.14).

$$\text{hrv} = \text{hr(max)} - \text{hr(min)} \quad (3.14)$$

This was symbolically represented based on the classifications of previous studies (Low et al, 1971; Krebs, 1979), shown in table 3.6.

Variability (bpm)	Classification
0 - 2	Absent
3 - 5	Reduced
6 - 25	Normal
> 25	Increased

Table 3.6: Classification of heart rate variability.

The heart rate variability was measured for all one minutes epochs free from accelerations and decelerations within a 15 minute segment of CTG. A histogram of the classifications was plotted (figure 3.23).

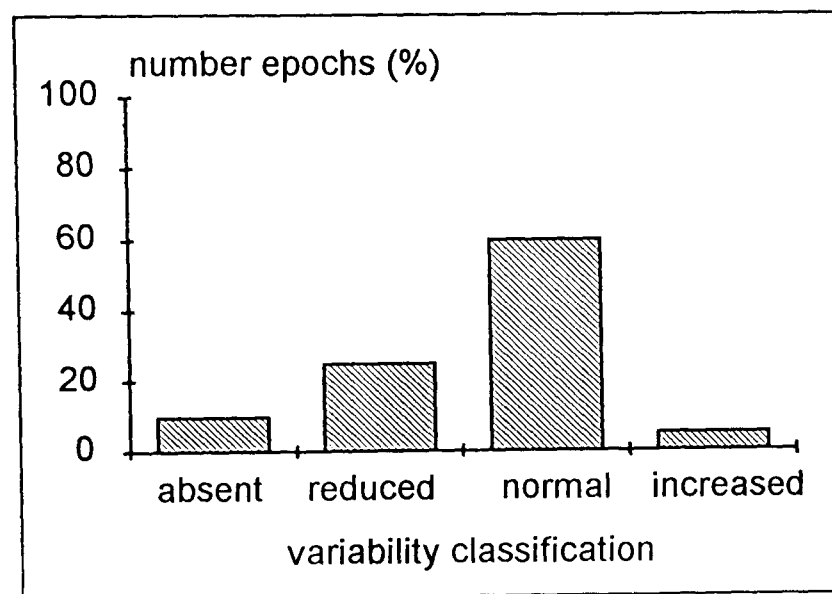


Figure 3.23: Histogram of variability classifications.

It was first believed that the overall classification of the histogram would be given by the heart rate variability for which the most epochs were classified (the modal classification). This interpretation was compared with an expert during a review of 10 labours but was found to be unsuitable. A more complex set of rules were required to interpret the histogram which were developed during a second review (figure 3.24).

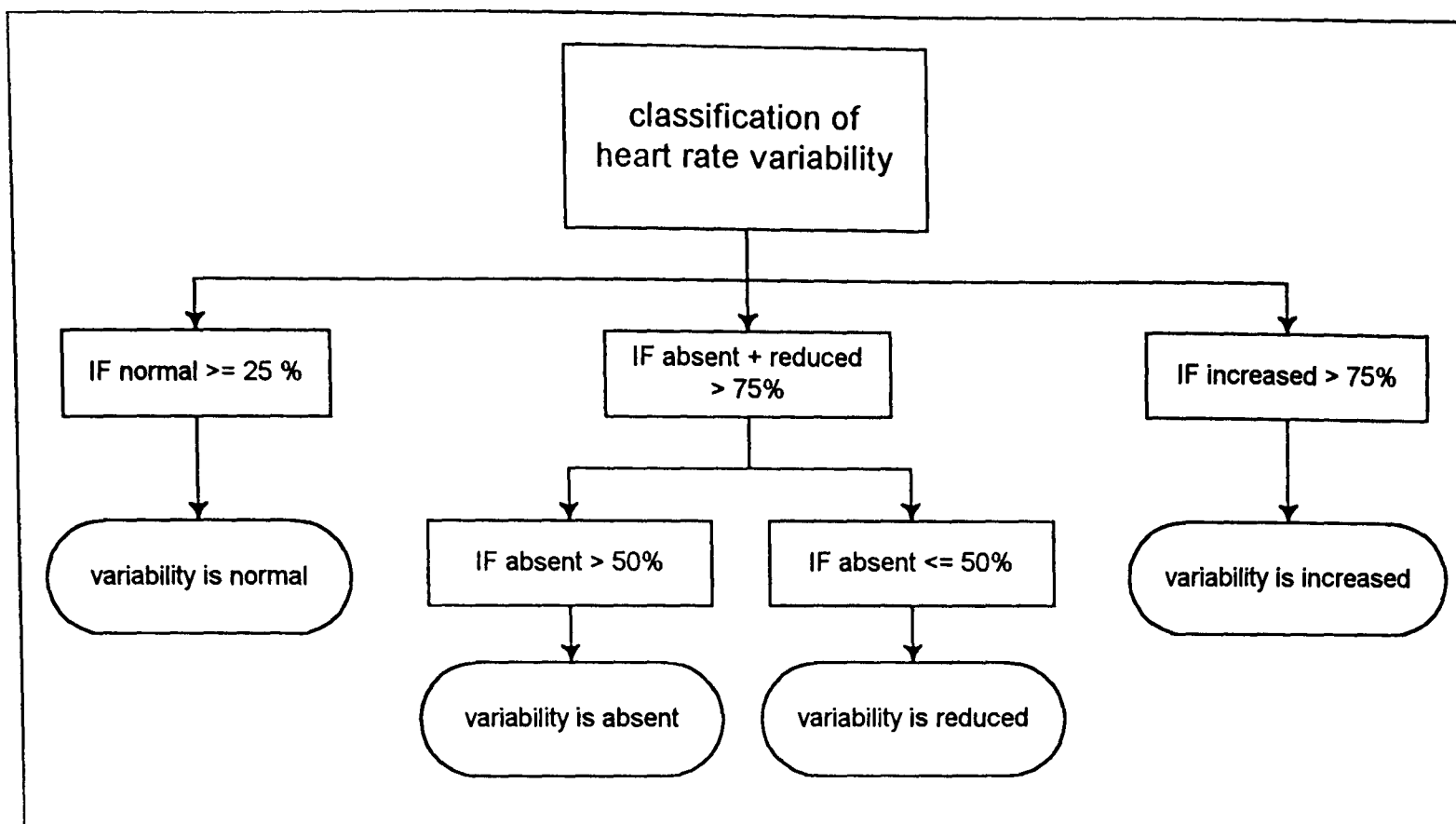


Figure 3.24: Rules to interpret the variability histogram.

It would be possible for a situation to arise where the histogram could not be classified using these rules; for example, if 20% of epochs were normal, 20% were absent, 20% were reduced and 40% were increased. This situation would be extremely unlikely as it requires both increased and absent variability to occur in the same 15 minutes of recording. However, if this did happen then the variability would be classified as inconclusive.

3.10 Contractions.

The location and magnitude of the contractions are important for classifying the relative timing of heart rate decelerations as well as for monitoring the administration of drugs used to augment labour.

Contractions were identified using a similar approach as that for obtaining the baseline heart rate and accelerations. Firstly a reference signal was established (figure 3.25).

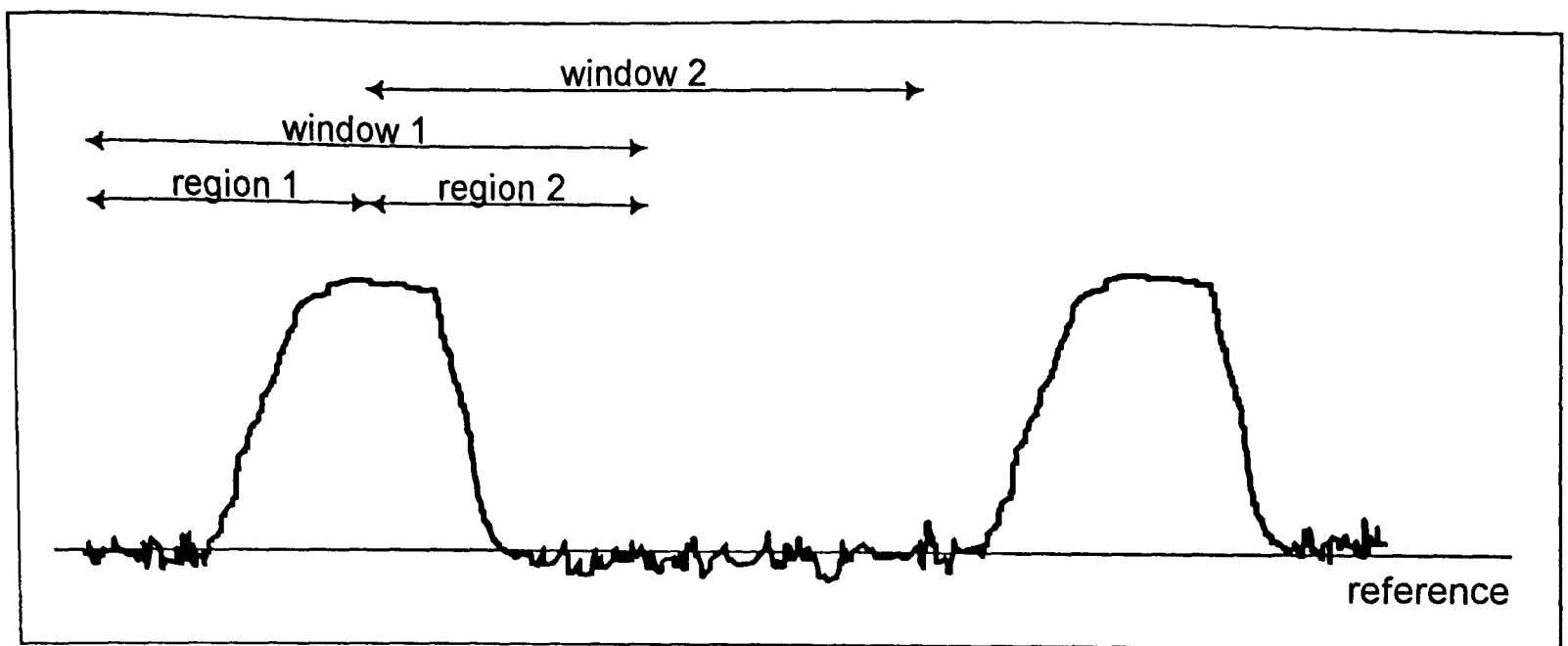


Figure 3.25: Fitting the contraction ground level.

The reference level for region 1, of length 1 minute, was obtained by representing the data samples in window 1 (2 minutes), in a statistical frequency distribution. The reference value was taken as the 5th percentile of this distribution. This, rather than the minimum value, was used to reduce the risk of noise samples being incorrectly taken as the reference signal. A similar process identified the reference level in region 2. The model used to classify contractions is shown in figure 3.26.

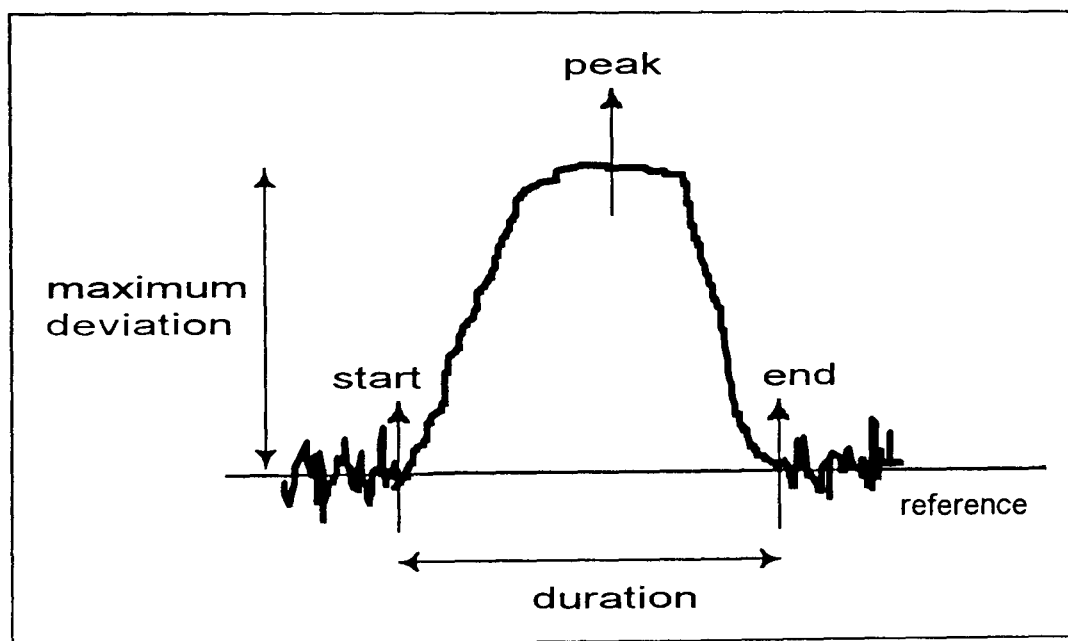


Figure 3.26: The contraction model.

The contraction samples had been digitised to 8 bit resolution (0 to 255). For a feature to be identified as a contraction, the maximum deviation above the reference had to be greater than 25,

$$x(i) > \text{reference} + 25$$

Where $x(i)$ is a contraction sample and the duration of the feature had to be longer than 20 seconds. The start of the contraction was recognised when, $x(i) > \text{reference} + 5$, for 3

seconds. The end of contraction was identified when $x(i) < \text{reference} + 5$, for 3 seconds. This period was chosen to ensure that the true beginning and end of the contraction had been identified as opposed to a transitory noise spike which may be sufficient to cross the reference line and thereby lead to incorrect detections. This figure is clearly arbitrary but was found to be suitable during evaluation.

The peak of the contraction was taken as the maximum data value between the start and end of the contraction. The algorithm is represented in greater detail in figure 3.27.

If a heart rate deceleration was detected, then the timing of this event was related to the contraction signal to determine whether the deceleration occurred simultaneously or lagged a contraction. If the deceleration's minima lagged the contraction peak by greater than 20 seconds then the deceleration was termed 'late', otherwise the deceleration was termed 'early or variable'.

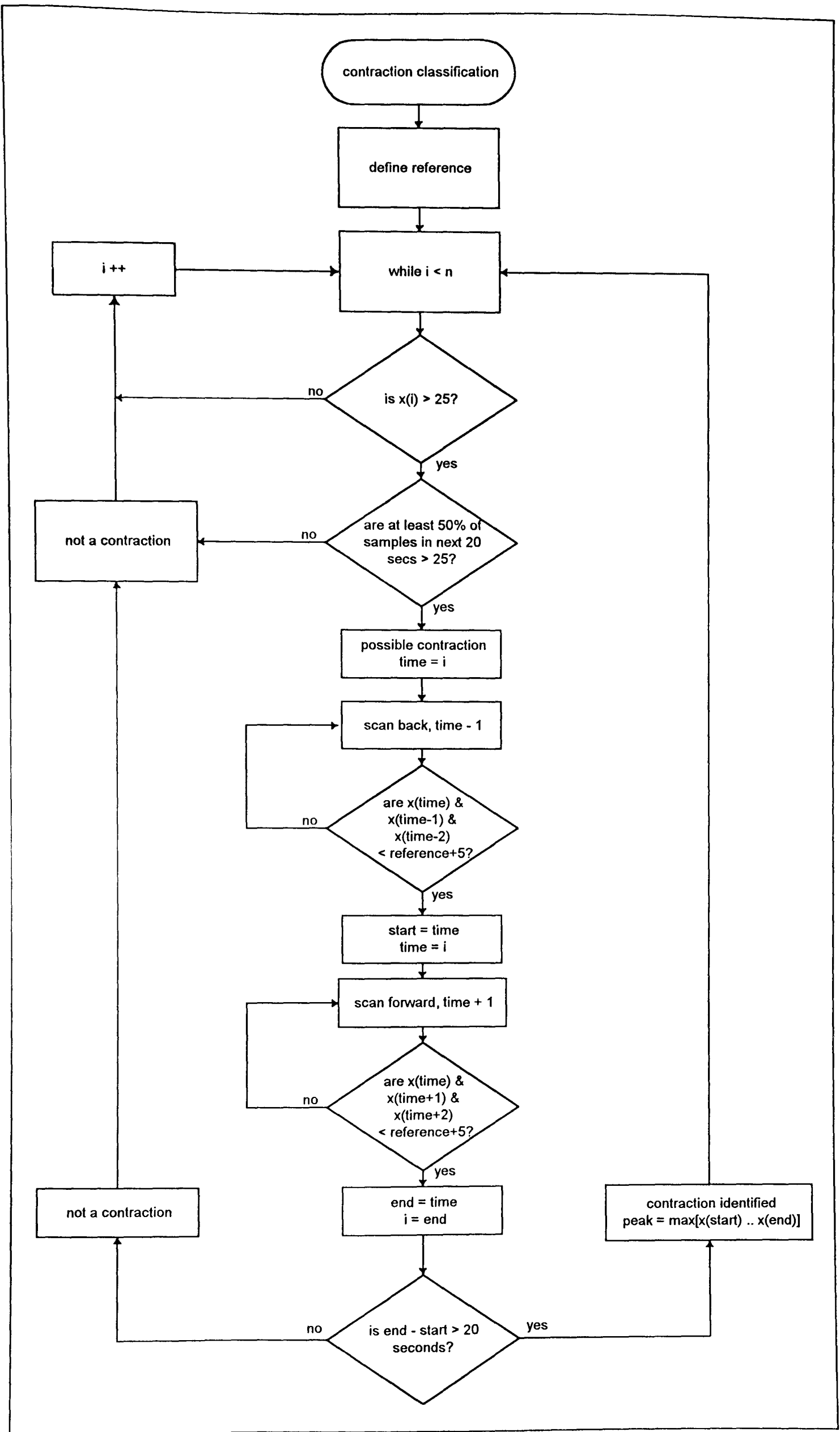


Figure 3.27: Contraction identification algorithm.

3.11 Signal quality.

3.11.1 Heart rate data.

The recording equipment recorded the time measured in milli-seconds, between successive heart beats. All measurements made during each minute of recording were stored in one minute data bins. If all heart beats had been detected, the sum of the timing measurements in a given data bin would equal one minute or 60,000 ms. This fact was used to obtain an indication of the heart rate signal quality. If the sum of these measures was, $> 54,000$ ms (90%), then the data bin was considered to have good quality data. If the sum of the measures was between 80% - 90%, then the data quality was considered adequate and if the sum of the measures was $< 80\%$ then the data quality was considered poor. These thresholds were clearly arbitrary, but they provided a reasonable assessment. A collective classification of the last 10 minutes of recording was made using the rules,

IF (all assessments in last 5 minutes were good)
THEN signal quality is GOOD.

IF (all assessments in last 5 minutes were adequate)
THEN signal quality is ADEQUATE

IF (all assessments in last 5 minutes were poor)
THEN signal quality is POOR

IF (number of good + number of adequate assessments in last 10 minutes > 8)
THEN signal quality is GOOD

IF (number of good + number of adequate assessments in last 10 minutes > 5 & ≤ 8)
THEN signal quality is ADEQUATE

IF (number of good + number of adequate assessments in last 10 minutes ≤ 5)
THEN signal quality is POOR

3.11.2 Contraction data.

It was found that when the contraction data was absent or of poor quality, the signal obtained was small and did not vary much above the reference level. The graphs in figure 3.28 show the likely statistical frequency distributions of the variations above the reference, δ , for good, adequate and poor quality data. Where,

$$\delta = \text{sample value} - \text{reference}$$

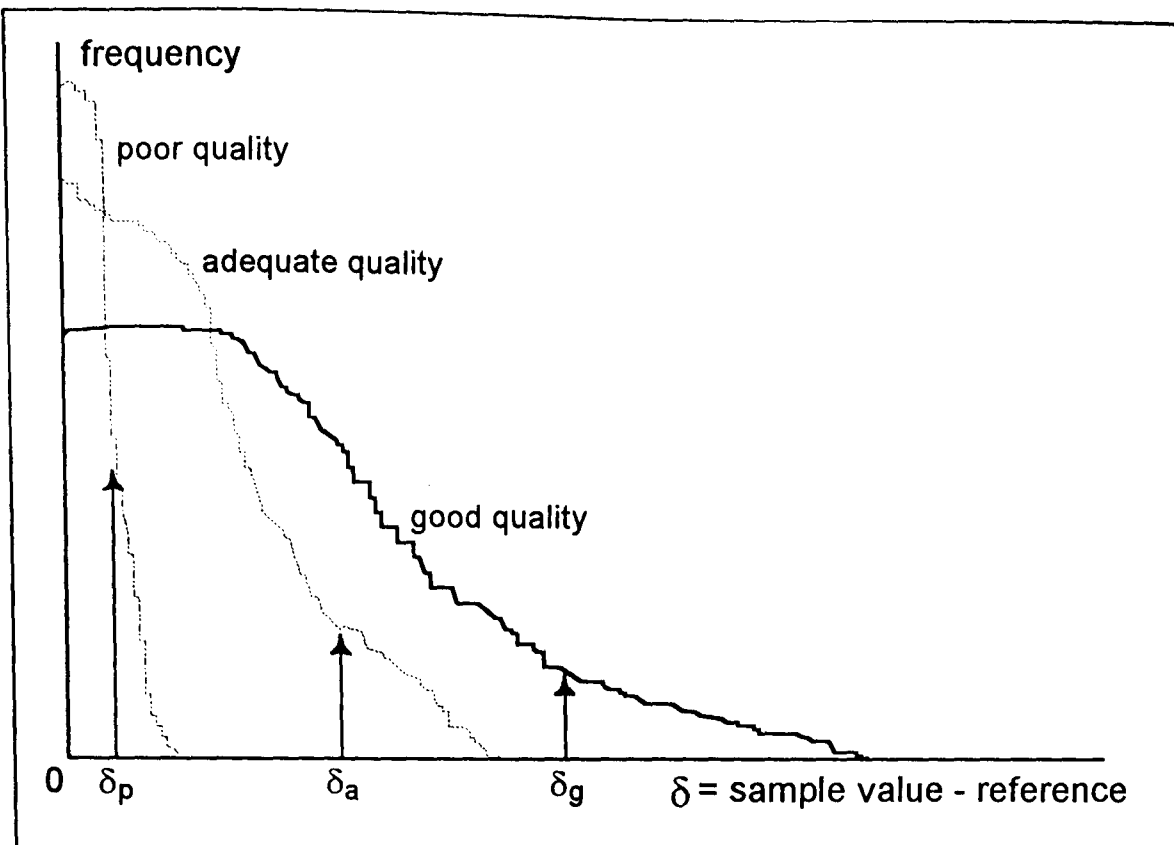


Figure 3.28: Statistical frequency distributions of δ ($\delta = \text{sample value} - \text{reference}$) for good, adequate and poor contraction signal quality.

The values, δ_p , δ_a , and δ_g , indicate the 90th percentile point of the distribution of δ . It was found that an indication of good quality data was obtained when $\delta_g > 40$, adequate quality when, $40 \geq \delta_a > 25$, and poor signal quality when $\delta_p \leq 25$. These were represented as a green, amber and red light respectively by the user interface. The flow diagram for this algorithm is shown in figure 3.29.

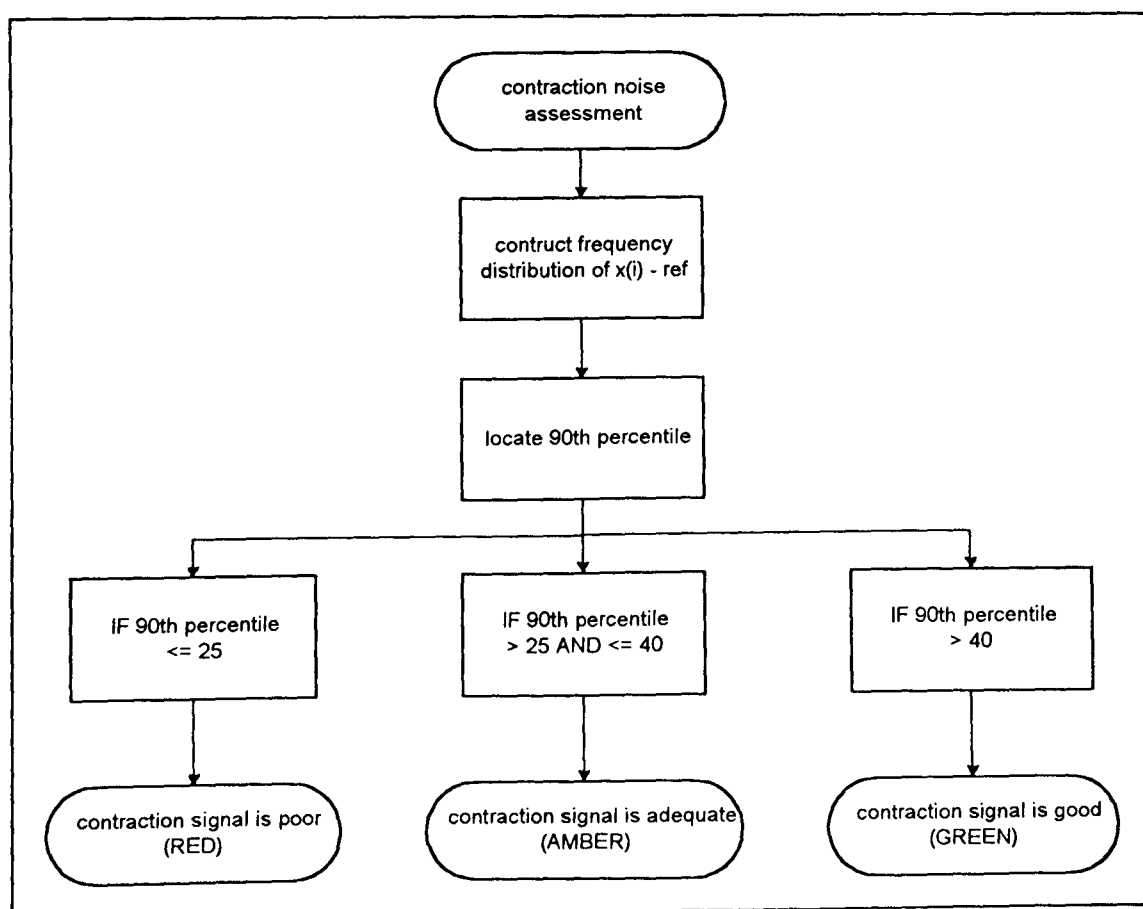


Figure 3.29: Rules to interpret the contraction signal quality.

Chapter 3 summary.

It was proposed to obtain the fetal heart rate and uterine contraction signals from a conventional intrapartum CTG recorder. These signals would pass through a process of feature extraction to obtain the important information. This, together with the information pertaining to the specific pregnancy and labour would then be processed by an expert system.

Chapter 3 describes the development of methods to extract the features from the cardiotocogram. These methods were later incorporated into the intelligent system. The important features were considered to be baseline heart rate, heart rate variability, accelerations, decelerations and the location of contractions. In addition, it was considered important to measure the signal quality of the heart rate and contraction data.

The performance of the system would depend on its ability to classify features from the CTG similarly to experts. An investigation was undertaken to assess the suitability of artificial neural networks (NNs) for feature extraction as they can be trained by experts to recognise patterns in data. This study found NNs suitable for feature extraction when the problem was reduced to small, well defined tasks and numerical algorithms were used to pre-process the raw data before it was applied to the NNs. A NN with optimised dimensions was used in this way to classify the magnitude of decelerations, a feature clinicians find particularly difficult. The NN was compared with six reviewers which included two CTG experts. The experts were consistent (89.7% and 97.0%) and agreed well with each other (81.0%) whereas the non-experts agreed less well. The NN was found to agree well with the experts (75.0% and 81.9%)

Earlier attempts to fully classify the raw CTG using a single NN were unsuccessful because of the large number of possible data patterns. A simplified approach to classify the magnitude and timing of decelerations was also unsuitable when contraction data was of poor quality or absent.

The extraction and classification of the remaining features was accomplished using numerical algorithms developed closely with experts. In each case these methods were found to compare well with experts.

Chapter 3 references.

- Allred L G and Kelly G E. (1990) Supervised learning techniques for backpropagation networks. Proceedings International Joint Conference on Neural Networks. July 17-21 San Diego. IEEE Neural Networks Council. 721-728.
- Dawes G S, Houghton C R S, Redman C W G and Visser G H A. (1982) Pattern of the normal fetal heart rate. *Br J Obstet Gynaecol.* **89**: 276-284.
- Dawes G S, Moulden M and Redman C W G. (1991) System 8000: Computerized antenatal FHR analysis. *J Perinat Med.* **19**: 47-51.
- Hon E H. (1968) An atlas of fetal heart rate patterns. Harty Press New Haven, CT. USA.
- Keith R D F, Westgate J, Ifeachor E C and Greene K R. (1993) Suitability of artificial neural networks for feature extraction from the cardiotocogram during labour. *Med Biol Eng & Comp.*(Accepted for publication)
- Krause W. (1990) Natalie by Niess. A computer-aided monitoring system for supervision of labour. In: Computers in perinatal medicine. Maeda K. (ed). Elsevier Science Publishers, B. V. Amsterdam, 103-111.
- Krebs H B, Petres R E, Dunn L J, Jordaan H V F and Segretti A (1979) Intrapartum heart rate monitoring I. Classification and prognosis of fetal heart rate patterns. *Am J Obstet Gynecol* **133**: 7, 762-780.
- Lindgaard O, Bottcher L M, and Weber T. (1992) Description, evaluation and clinical decision making according to various fetal heart rate patterns. Inter-observer and regional variability. *Acta Obstet Gynecol Scand* **71**: 48-53.
- Low J A, Boston R W and Pancham S R. (1971) The role of fetal heart rate patterns in the recognition of fetal asphyxia with metabolic acidosis. *Am J Obstet Gynecol.* **109**: 922.
- Maeda K. (1990) Computerised analysis of cardiotocograms and fetal movements. Balliere's Clinical Obstetrics and Gynaecology. Lilford R (ed) **4**: 797-81.
- Rumelhart D E, Hinton G E, and Williams R J. (1986) Learning Internal Representations by Error Propagation. In: Parallel Distributed Processing Feldman J A, Hayes P J and Rumelhart D E. (eds.), The MIT Press, Cambridge England. **1**: 318 - 362.

Chapter 4

Development of an intelligent system for labour management:

The expert system.

4.1	Introduction	101
4.2	Knowledge elicitation	102
4.3	Knowledge representation	104
4.4	System implementation	107
4.5	System integration	115
	Chapter 4 summary	118
	Chapter 4 references	119

4.1 Introduction.

The development of the expert system was based on the general model shown in figure 4.1.

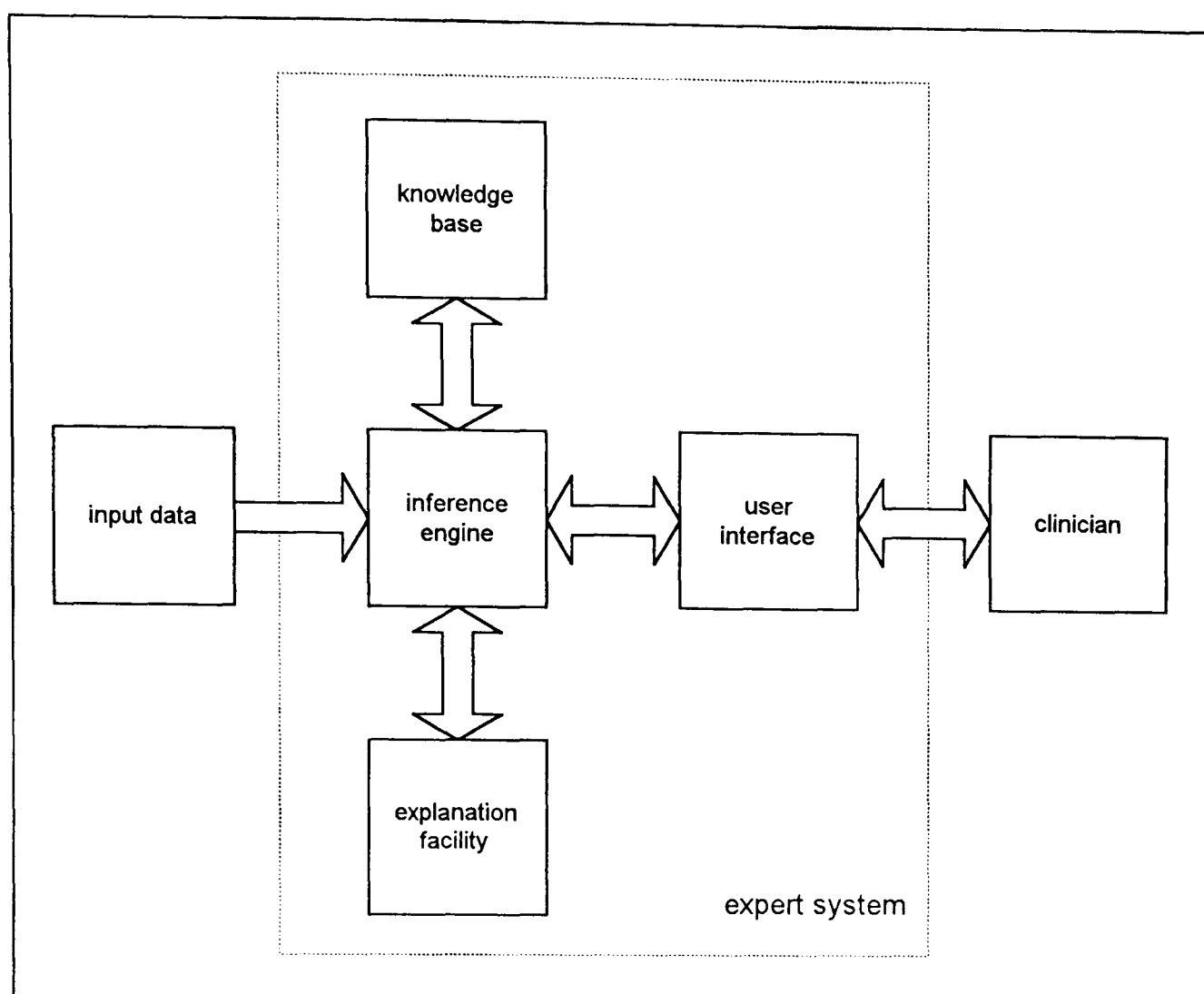


Figure 4.1: The expert system.

The features extracted from the CTG discussed in chapter 3, together with the relevant clinical information, form the input data to the expert system. This information is known as the short term or dynamic knowledge, which is processed by the inference engine. The inference engine reasons with this information by applying long term knowledge represented as rules in the knowledge base to infer facts regarding the input data. These facts ultimately lead to conclusions or recommended actions which are then communicated to the user in a meaningful way via the user interface. If the inference engine requires further information in order to prove certain rules, then it can ask questions of the user. Alternatively, if the user wished to know how certain conclusions were reached, then the explanation facility could elucidate on the rules which were proved true and led to the recommended action.

There were 5 stages to the development of the expert system.

1. Knowledge elicitation.
2. Knowledge representation.
3. Expert system implementation.
4. Expert system evaluation and combined system evaluation.
5. Combined system validation.

The evaluation and validation examined the performance of the system at the key stages in its development. These are discussed in chapter 5.

4.2 Knowledge elicitation.

The acquisition of knowledge, if handled inappropriately, can be a major constraint in the development of expert systems (Cullen and Bryman, 1988). Methods for knowledge elicitation are continually being developed and reviewed (Neale, 1988) and practical strategies have been well documented (Boose, 1984; Davies and Hakiel, 1988; Diaper, 1989). The two strategies which have emerged as the most useful for developing expert systems are, the *fast prototyping* and the *evolutionary methods* (Hayward, 1987). Fast prototyping seeks to acquire knowledge from experts during generally unstructured interviews. This leads to the rapid development of a working, if limited, expert system. The prototype system can then be used to demonstrate the knowledge base to the experts and thereby act as a focus for discussion. This would stimulate the experts to recommend further refinements to the existing knowledge and also identify areas which required extending. The evolutionary method involves developing the expert system over a longer period of time. For this strategy, the knowledge is represented in a non-system form such as logic diagrams, tree diagrams, or semantic nets to stimulate feedback.

For our application, it was decided to use a combination of these methods. In the first place, knowledge elicitation was carried out with two experienced obstetricians in relatively unstructured, interactive group sessions. All members of the group knew one another well and did not feel inhibited (or intimidated!) which was considered an advantage. It has been found that this type of approach generates more accurate data with a greater quantity of ideas than other approaches. This has been attributed to the synergism created by the experts sharing their thoughts (Meyer and Booker, 1991). It has also been shown from a comparison of elicitation methods, that the interactive group method produces a greater quantity of ideas and higher member satisfaction than any other approach (Seaver, 1976). Each session was conducted, where possible, with flexible time constraints and each member present (including the engineers) was free to contribute at will. All sessions were

tape recorded and a full transcript was later produced. Four questions were considered during these sessions which were given to the experts prior to the first meeting.

- Question 1: Ignoring the possible constraints of technology, what conceptually do you wish to see at the end of this project? For example,
- How and what information should pass from the system to the user?
 - What role should it play in managing labour?
 - Who would use the system?
- Question 2: Convey in outline form only, the procedure you adopt in managing labour. For example,
- When is fetal heart rate monitoring started?
 - How often is the CTG assessed and by whom?
 - Who are the people involved in the management of labour?
- Question 3: Write down in any convenient form (such as rules, decision trees, diagrams or sketches) how you interpret the CTG. How often should this be done?
- Question 4: Select several CTG records which represent various categories of patients. Highlight the sections of significance and explain why you consider them important.

The first session discussed questions 1 and 2 and led to the conceptual model discussed in chapter 3 (section 3.1, figure 3.1). Five further sessions which lasted a total of 15 hours took place during which questions 3 and 4 were discussed. The transcript produced in session 2 is shown in appendix C. At the end of these meetings it was considered that sufficient information had been obtained to allow the essential aspects of the knowledge to be formalised and implemented in a small expert system based on the fast prototype model. This expert system contained approximately 150 rules and was written in Prolog software and is discussed in 4.4. It was decided to develop this limited system to confirm that the essentials of labour management could be accurately represented and to serve as a stimulus for the experts to indicate areas they considered required modification. An evaluation of this system was undertaken (Ifeachor et al, 1991) and is discussed in chapter 5 (section 5.3). This evaluation demonstrated the feasibility of an expert system for labour management which allowed an evolutionary model to then be adopted. Over a period of two years the system was extended and modified in line with expert recommendations until it was considered ready to be compared with independent experts.

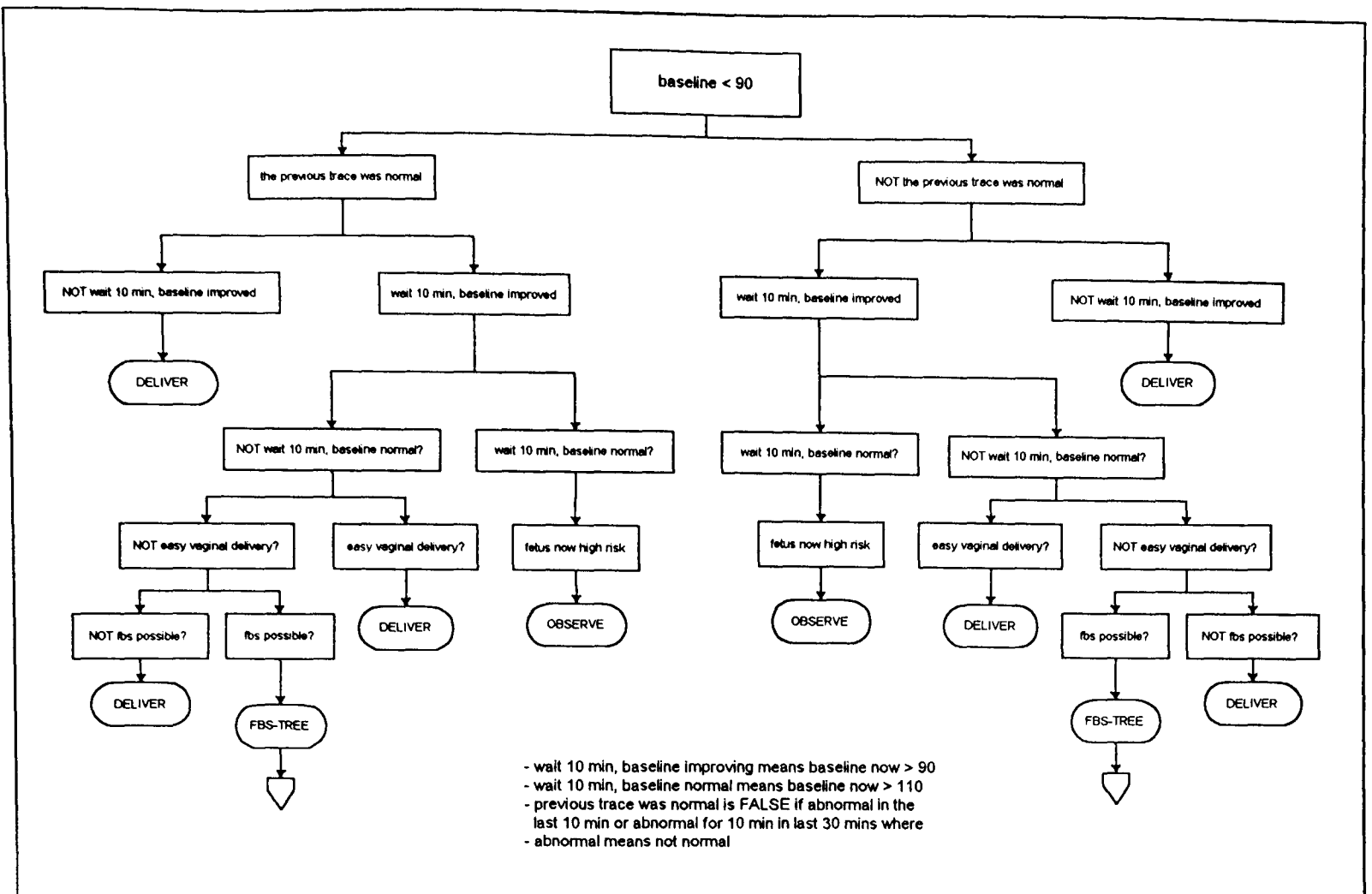


Figure 4.3: Baseline less than 90 bpm knowledge tree.

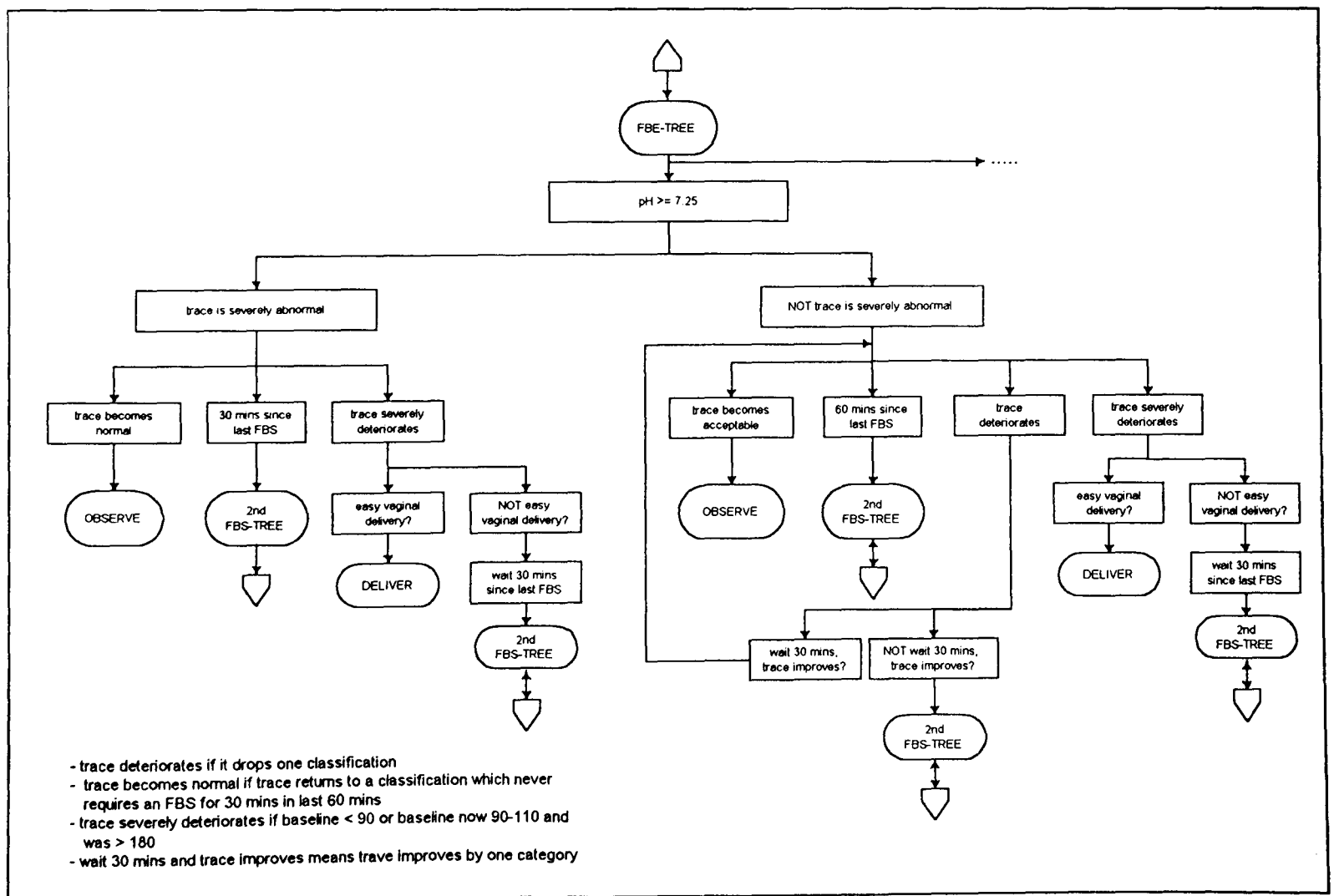


Figure 4.4: Fetal blood sample tree (a); pH ≥ 7.25.

4.3 Knowledge representation.

The most useful representation of the knowledge was found to be in the form of decision tree diagrams. These provided a structure for the knowledge as well as the means for the experts and the engineer to readily discuss specific aspects of the knowledge and suggest refinements.

The current state of the knowledge is represented in appendix D and selected branches are shown in figures 4.3, 4.4 and 4.5. Figure 4.3 shows the knowledge required to manage labour when the baseline fetal heart rate falls below 90 bpm, which is a serious event. Figures 4.4 and 4.5, shows how fetal blood sampling has been incorporated into the system's decision making. The key to interpreting the knowledge tree is given in figure 4.2. Each branch begins with an assessment of the CTG and ends with a conclusion or recommended action.

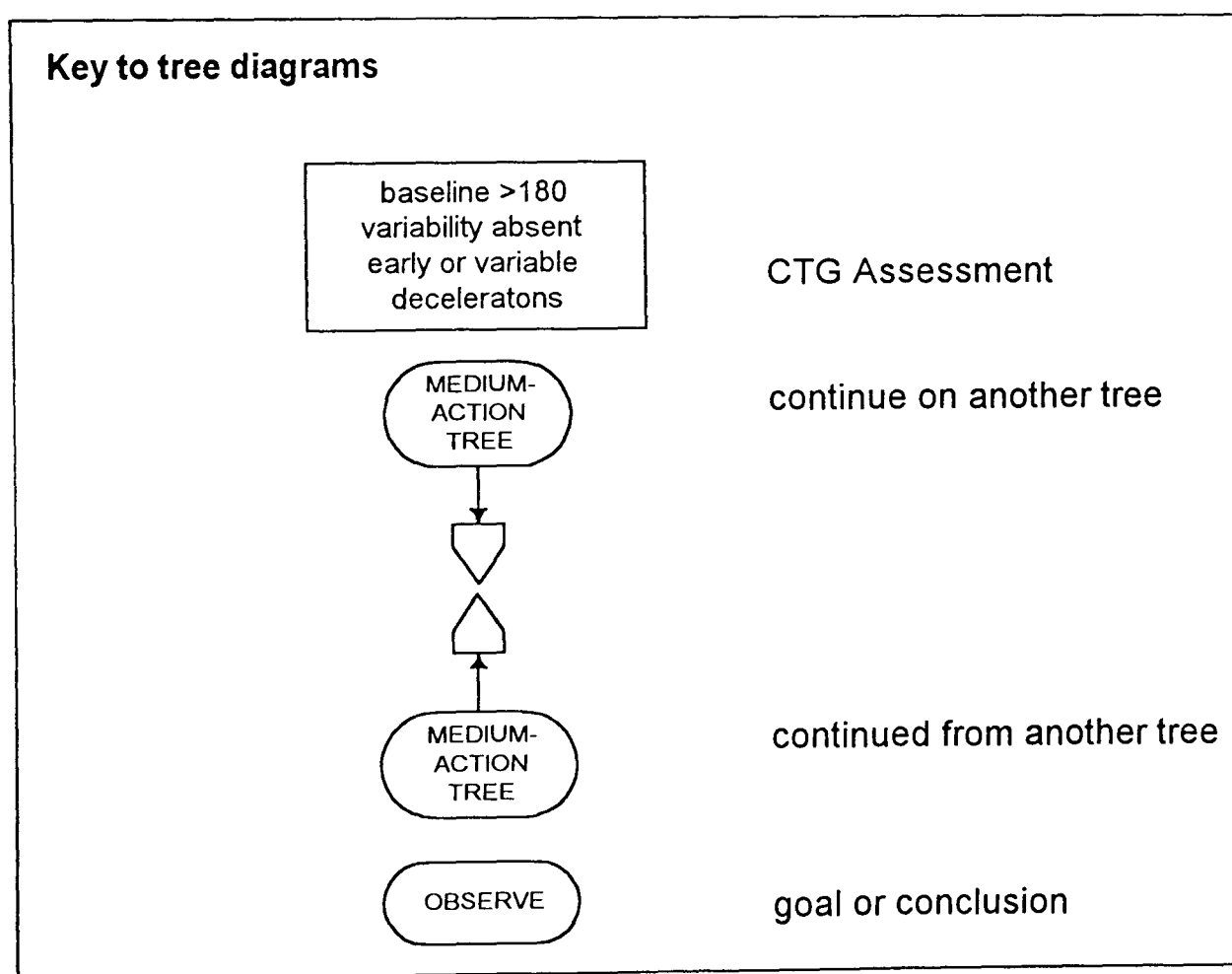


Figure 4.2: Key to knowledge tree.

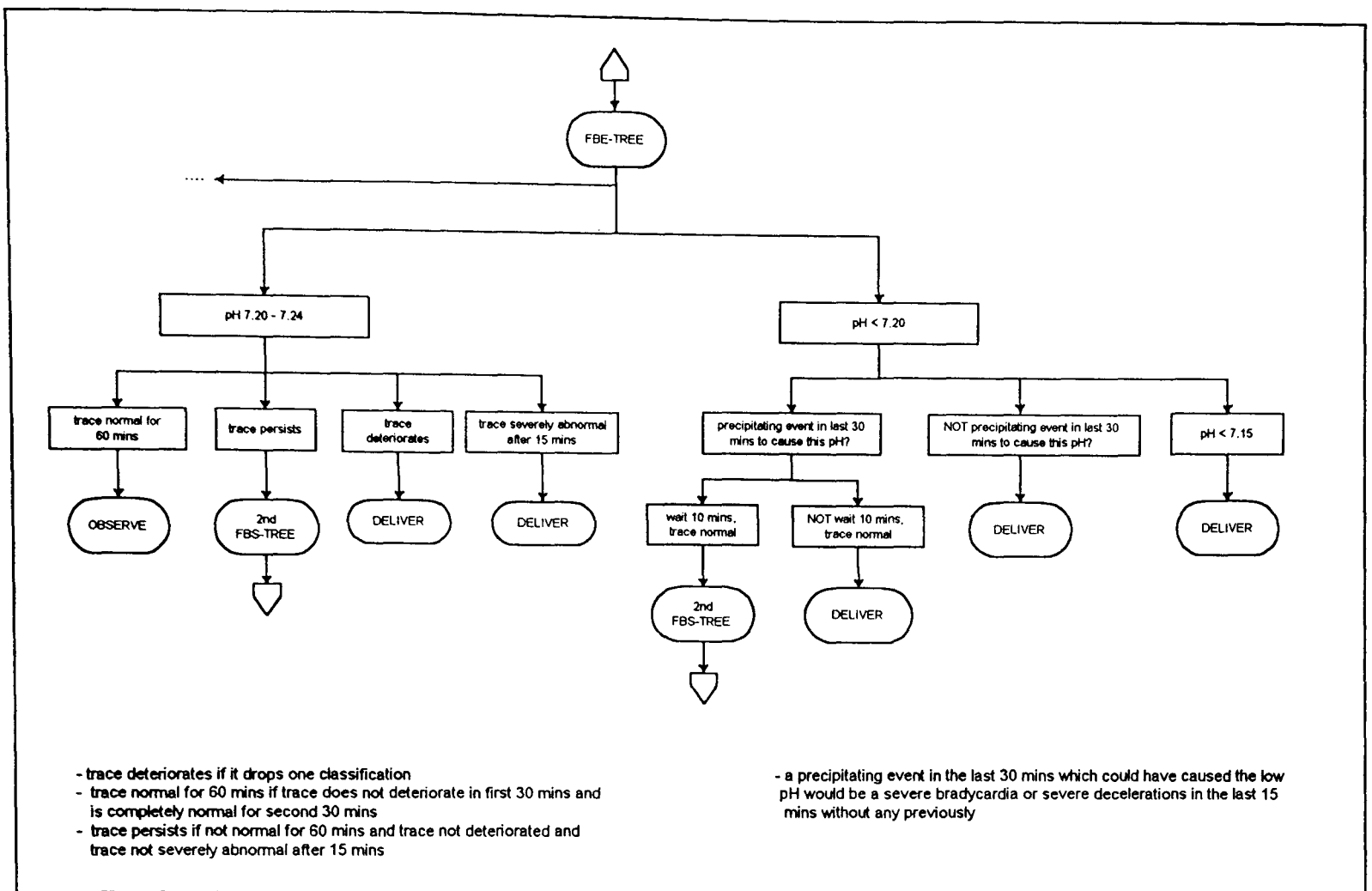


Figure 4.5: Fetal blood sample tree (b); pH < 7.25.

4.4 System implementation.

4.4.1 The knowledge base.

The knowledge described by the decision trees was represented in the form of production rules which took the following form.

Rule(*rule number, head, tail, [conditions]*)

The rule number identified the rule. The rule head indicated the current position in the knowledge tree and the rule tail indicated the position in the tree to be moved to if the associated conditions were proved true. If a condition was proved false, then the rule was false and an alternative rule with the same head was searched for. When a rule was proved true, progress had been made towards a goal, whereupon the rule tail became the new head and a new rule was searched for with this head. This procedure continued until a goal at the end of a branch was found. To illustrate this process, consider the following example which attempts to identify an animal based on its description.

Some facts regarding animals are,

1. IF an animal,
 is warm blooded AND suckles its young AND has hair,
 THEN the animal is a mammal.
2. IF a mammal,
 swims in the sea AND is the largest creature on Earth,
 THEN the mammal is a blue whale.
3. IF a mammal,
 walks on two legs AND drives a car,
 THEN the mammal is a human.
4. IF a human,
 has babies,
 THEN the human is a woman.

These facts can be written concisely as the following production rules,

Rule(1, animal, mammal, [is warm blooded],[suckles its young], [has hair])

Rule(2, mammal, blue whale, [swims in sea],[largest creature on Earth])

Rule(3, mammal, human, [walks on two legs],[drives a car])

Rule(4, human, woman, [has babies])

In the same way, each position in the labour management tree was assigned a unique identifier. The principal node at the foot of the tree was identified as 'heart_rate'. Consider the tree in figure 4.3 which considers the management required when the baseline heart rate falls below 90 bpm.

Some facts which relate to this are,

1. IF heart_rate,
 is less than 90 bpm,
 THEN baseline < 90.
2. IF baseline < 90,
 the previous trace was normal AND NOT after 10 mins the baseline has improved,
 THEN DELIVER.

which can be written concisely as,

```
rule(1, heart rate, baseline < 90, [1])  
rule(2, baseline < 90, Deliver, [2][3])
```

```
cond(1, the baseline is < 90 bpm)  
cond(2, the previous trace was normal)  
cond(3, NOT after 10 mins the baseline has improved)
```

The rules obtained for the entire knowledge base were obtained in this way and can be found in appendix E.

4.4.2 Development of the inference engine.

The inference engine is the heart, or perhaps more precisely, the brain of an expert system. Its role is to prove rules and establish facts concerning the dynamic knowledge using the long term knowledge formed in the knowledge base. Its ultimate goal is to arrive at the end of a branch where an action can be recommended.

The expert system was first implemented as a fast prototype model written in Prolog software. The source code can be found in appendix F. This language was chosen because it has been designed to allow expert systems to be developed simply and quickly. However, it was found that Prolog had some serious draw-backs which, it was felt, would be difficult to overcome. Prolog is a high level language which can efficiently manipulate symbolic knowledge. Its limitations are that it provides few low level operations such as memory management, mathematical functions and data manipulation. These operations would be

essential in the full system which incorporated automatic feature extraction as these require a low-level language such as 'C'. A possible solution to the problem would have been to develop the expert system in Prolog and the feature extraction algorithms in 'C' and interface the two. However, it was considered that this approach would be difficult to manage and would not be sufficiently flexible to allow free movement of data. A second possible solution involved commercial software in the form of expert system shells which have been available for some time. These contain the fully developed constituent parts of the expert system except for the rule base which is developed for the specific application. This approach can obtain a usable system very quickly. The expert system shells available at the time were reviewed but no shell was found to be flexible enough to allow the simple integration of the feature extraction methods.

It was decided that the development of the system could best be achieved if the feature extraction algorithms and the expert system were both written in the C-programming language. This would allow the two main parts of the system to be merged into one software package to give maximum flexibility, low level control and efficient data flow. This approach did mean that an inference engine would need to be developed in a language that was perhaps not best suited for the task, but this approach was considered the best compromise. The developed inference engine algorithm is shown in figure 4.30 and the software developed in 'C' is given in appendix G.

The signal quality of the heart rate data had to be at least adequate before the inference engine was allowed to proceed. If the quality was unsatisfactory then the inference engine concluded that the quality of signal was too poor for monitoring purposes and suggested that the scalp electrode or ultra sound transducer be reapplied. If just the contraction signal quality was poor, then the expert system was still able to make some limited inferences regarding the CTG.

It will be recalled that a rule was represented as; *rule(rule number, head, tail, [conditions])* where the head indicated the current position in the knowledge tree and the tail pointed to the next position to move to if the conditions were proved true. When the inference engine was invoked, the current head was set to 'heart_rate'. This is starting position for the inference engine algorithm (figure 4.6) and represents the foundation upon which all branches were connected. Some facts would have already been established prior to running the inference engine, for example those conditions associated with feature extraction. If any condition had been proved *true* then the associated condition number would be stored in the, YES[], array. Similarly, if a condition had been proved *false*, then its condition number was stored in the, NO[], array. With reference to figure 4.6, the inference engine algorithm then continued,

1. The knowledge base was searched and a rule was obtained with a head the same as the current head. The rule was then split into its constituent parts; rule number, head, tail and dependent conditions.
2. The, NO[], array was checked to see if any of these conditions had been previously proved *false*, in which case the current rule could not be *true* and was abandoned in favour of another rule with the same current head. This procedure continued until a rule was found which had not been previously proved *false*. If all rules with the current head were *false*, then no solution was possible which indicated a gap in the knowledge had been discovered. This was useful during testing.
3. If none of the conditions were to be found in the, NO[], array then the procedure next checked to see if any conditions were known to be *true* by examining the, YES[], array.
4. If a condition was not in YES[], nor in NO[], then the condition must be proved. Some conditions could be proved by the inference engine automatically, others would need to be proved for the inference engine by a clinician. A database was developed which contained a list of all the conditions the inference engine could prove itself. Associated with each entry was the software required to obtain the proof. As an example consider, cond(67, the previous trace was normal). The inference engine would attempt to prove this condition by executing the software associated with condition 67, which examined the classifications made for previous segments of recording. If these were normal then the condition was *true*, otherwise it was *false*.
5. If the condition was not listed in the database then it indicated that its proof could not be obtained automatically and must be sought from the user. An example of such a condition would be, cond(55, is an easy forceps delivery possible?)
6. If the condition was proved *false* here, then the algorithm returned to step 1 where a new rule was found with the same current head.
7. If all conditions were found to be *true* then the rule must be *true*, in which case progress had been made towards a goal, whereupon the tail became the new current head and the algorithm returned to step 1.
8. If at least one rule had been proved *true* and no rules existed in the knowledge base with the new current head, then a goal had been reached. These goal states were represented in the rule tail as 'goal_X' where X was the goal number. Each goal number had a specific message which was revealed in a pop-up window on the computer screen for example,

Goal 34. There is reduced variability with no accelerations. However, this has not persisted long enough to warrant further action.

I recommend you continue to observe this case.

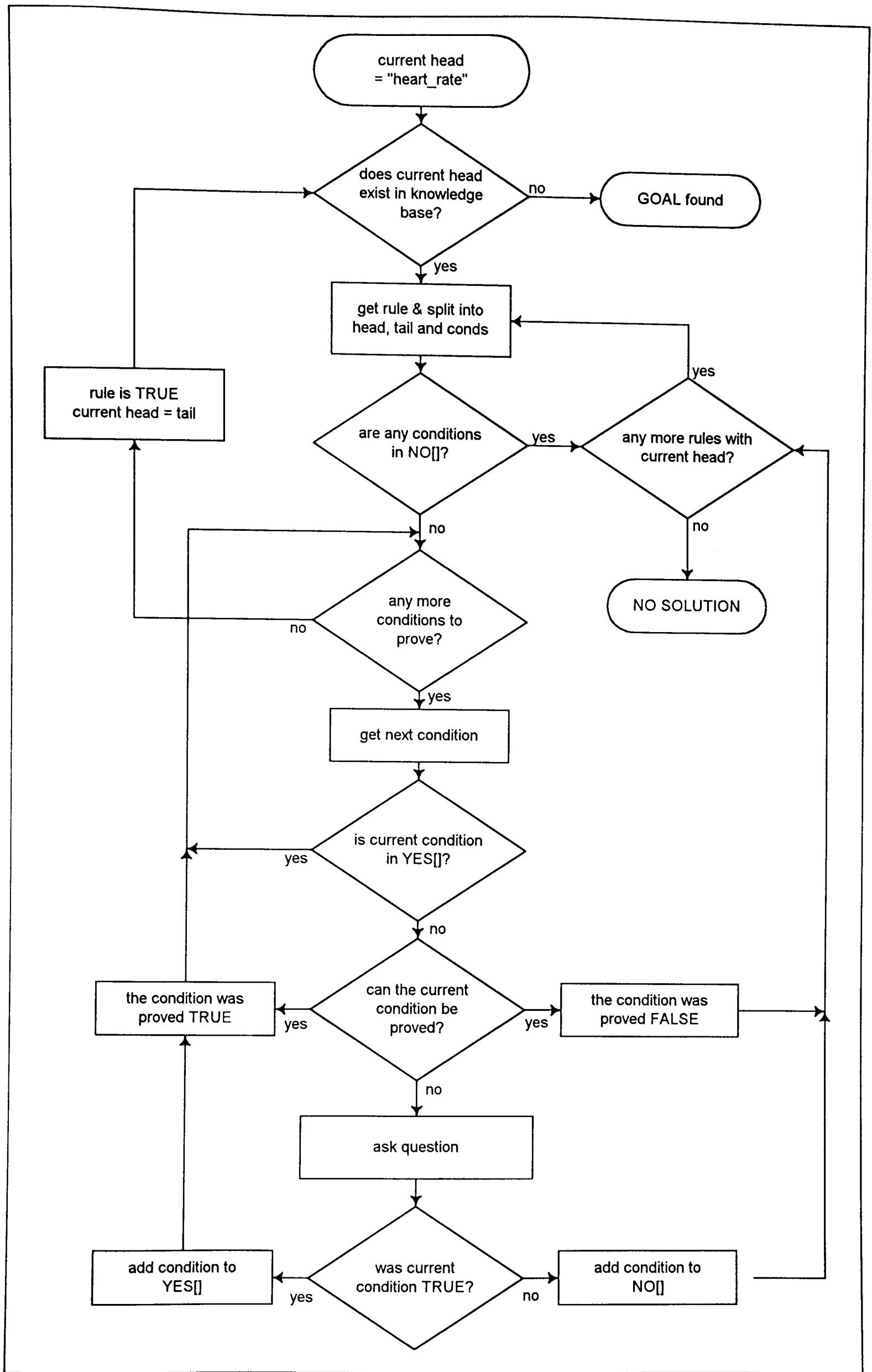


Figure 4.6: The inference engine.

4.4.3 The explanation facility.

The explanation facility was implemented in a limited form. It identified the rules and conditions which had been proved true and had led to the given action being recommended. The rule number was given with the associated condition numbers and a detailed explanation of what each condition was. This allowed the path taken by the expert system through the tree to be tracked and was also useful for testing.

4.4.4 The user interface.

The current system is a prototype system which has been implemented on a personal computer. The user interface has been designed to be informative to those involved in the system's development but is not adequate for use by clinical staff. All data entry was achieved using the keyboard. The system's output was displayed on the computer screen and presented the results from each feature extraction algorithm as well as messages from the expert system. When the expert system had reached a goal or conclusion, the associated message was displayed as a pop-up window. The output display is shown in figure 4.7.

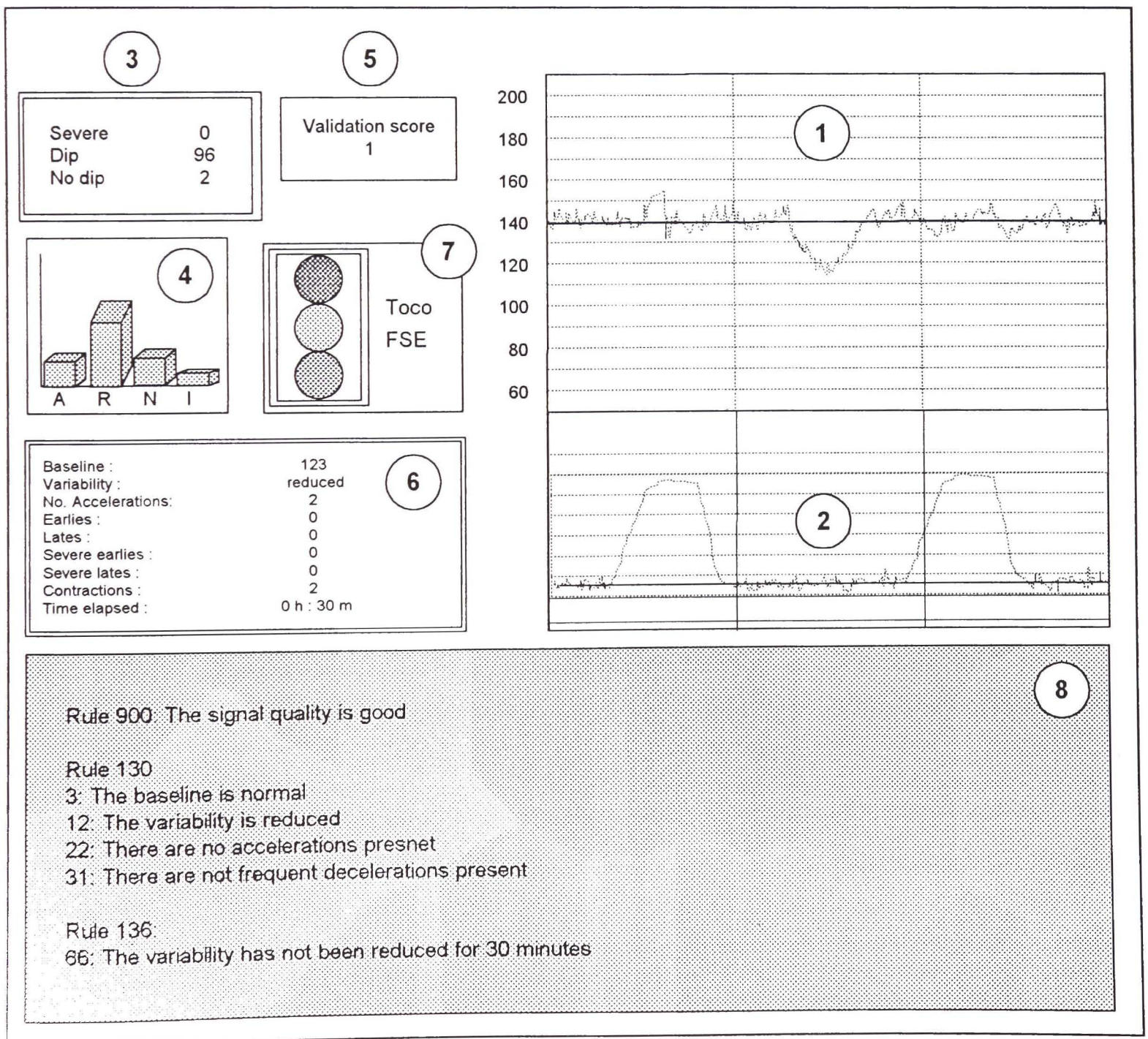
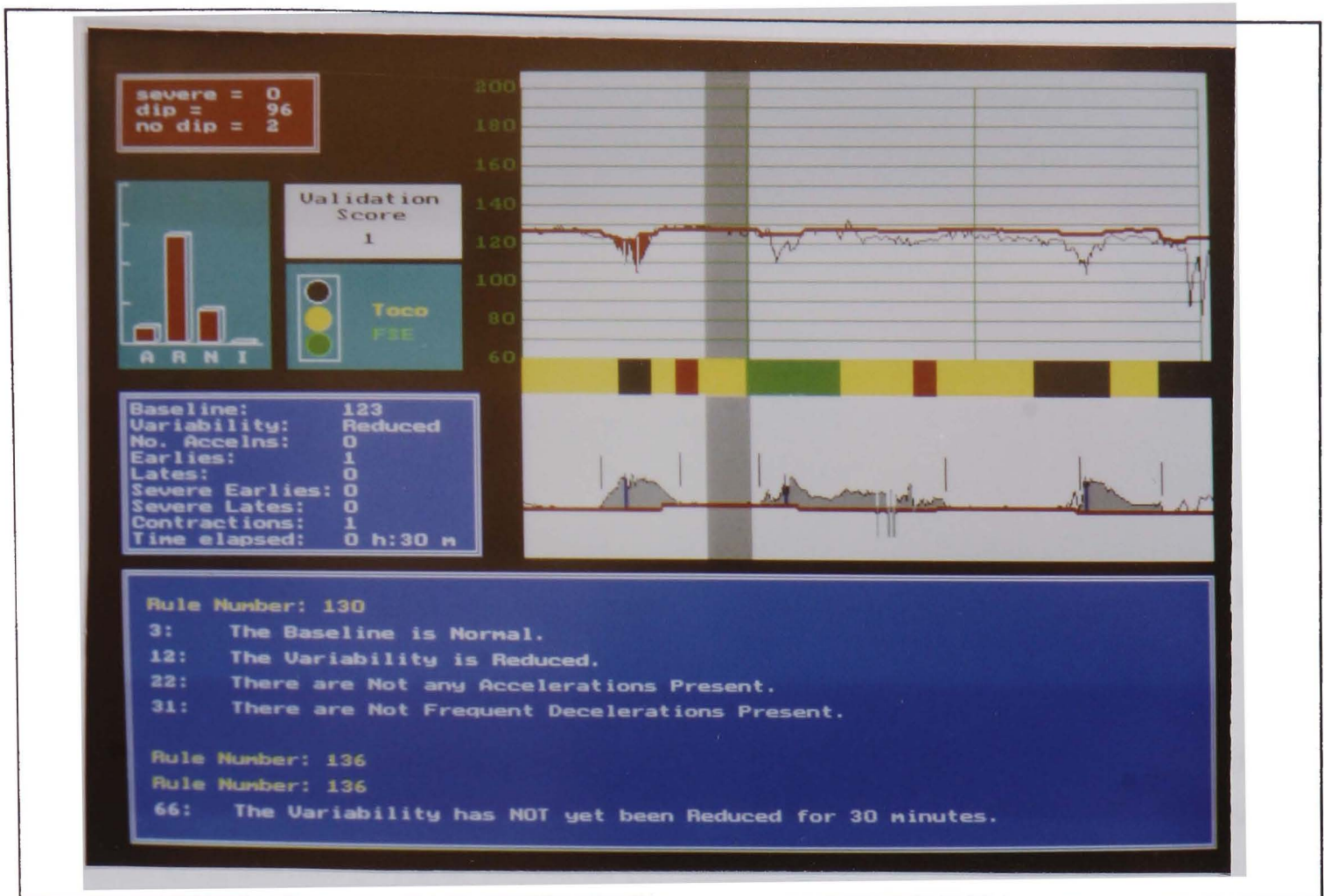


Figure 4.7: The user interface.

Key:

1. Heart rate data; the most recent 15 minutes of heart rate recording was plotted with the calculated baseline shown and any features classified as accelerations or decelerations indicated.
2. Contraction data; this was plotted with the calculated reference level shown and the start, peak and end of each detected contraction indicated.
3. Neural network output; the neural network classified the magnitude of decelerations as *severe*, *deceleration (dip)*, and *not deceleration (not dip)*. Each output ranged from 0 - 100% where an output value $> 80\%$ indicated a successful classification.
4. Variability classification; the histogram used for the heart rate variability calculation was plotted to indicate the percentage number of 1 minute epochs which had absent (A), reduced (R), normal (N) and increased (I) variability.
5. Validation protocol score; this number was printed for every 15 minutes of recording and indicated the concern the system had for the fetus. This score was used during the validation of the system, discussed in Chapter 5 (section 5.3).
6. CTG assessment; an assessment of the CTG together with the current record time was displayed and updated for every 5 minutes of recording .
7. Signal quality; the quality of the heart rate and contraction data was expressed in a traffic light display where a red light indicated poor signal quality.
8. The expert system; the explanation facility was the output of the expert system which detailed the rules and conditions which had led to a recommendation.

4.5 System integration.

The system was developed in two modules; the feature extraction module and the expert system. These were designed to be easily integrated to form the combined intelligent system as represented in figure 4.8.

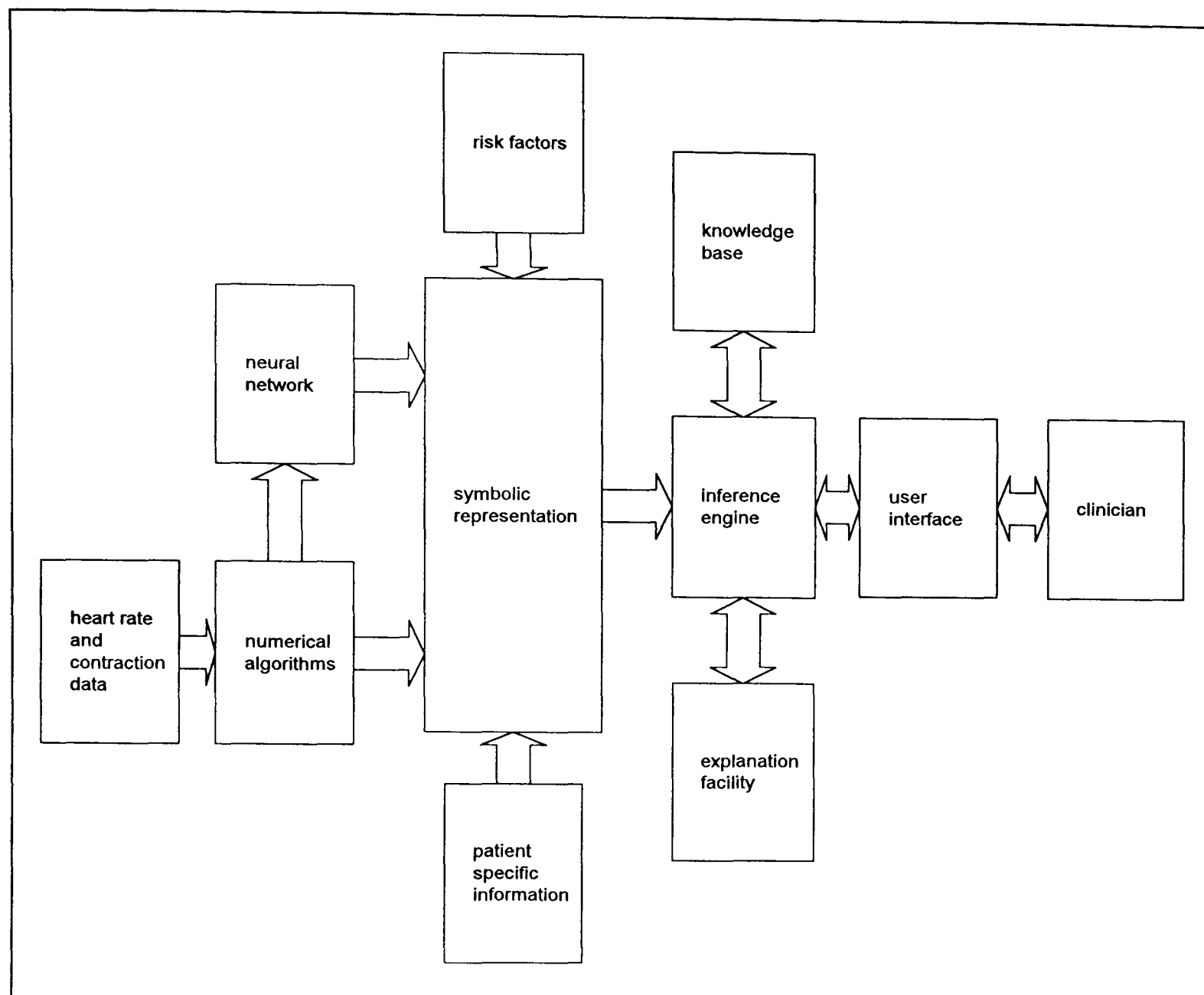


Figure 4.8: The complete system diagram.

This is a hybrid system which utilises the strengths of various techniques used in applied artificial intelligence. The important features were extracted from the CTG using numerical algorithms which were classified using additional algorithms and a neural network. These features, together with the patient specific information (mother and fetus), formed the dynamic knowledge which was symbolically represented using obstetric terminology (e.g. reduced variability, late decelerations etc.). This was then processed by an expert system which applied long term knowledge in the form of production rules represented in the knowledge base to infer facts and relationships which ultimately lead to actions being recommended.

4.5.1 Software control.

The system software was written in the C-programming language using the modular form represented in figure 4.9 and was implemented on an IBM compatible, Intel 80486DX microprocessor based personal computer. The system manager module controlled the flow of data from data collection, data processing through to data output. The other modules were independent of each other but shared common data which, for each 15 minutes of recording included;

1. Heart rate data.
2. Contraction data.
3. Baseline heart rate waveform. The calculated waveform about which accelerations and decelerations were referenced.
4. Deceleration map. The location of decelerations relative to the 15 minute window.
7. Acceleration map. The location of accelerations relative to the 15 minute window.
8. CTG assessment. An assessment of the last 5 minutes of recording. This was a global structure array which took the form,

```
typedef struct
{
    int    baseline_heart_rate;
    char   heart_rate_variability;
    int    number_accelerations;
    int    number_early_mild_decs;
    int    number_late_severe_decs;
    int    number_contractions;
    int    contraction_noise_assessment;
    int    heart_rate_noise_assessment;
    char   CTG_classification;
}
features;

features segment[240];
```

Where '*segment*' is an array of 240 '*features*' structures, which can hold an assessment of 20 hours of recording.

This combined system underwent a preliminary evaluation by comparing its recommended management of 30 labours with 3 experts from Plymouth (Keith et al, 1993) and is described in chapter 5 (section 5.3). This study found the complete system obtained a performance comparable with the experts and was therefore ready for a rigorous external validation. This study compared the system with 17 experts from 16 leading centres in fetal monitoring around the UK and is described in chapter 5 (section 5.4).

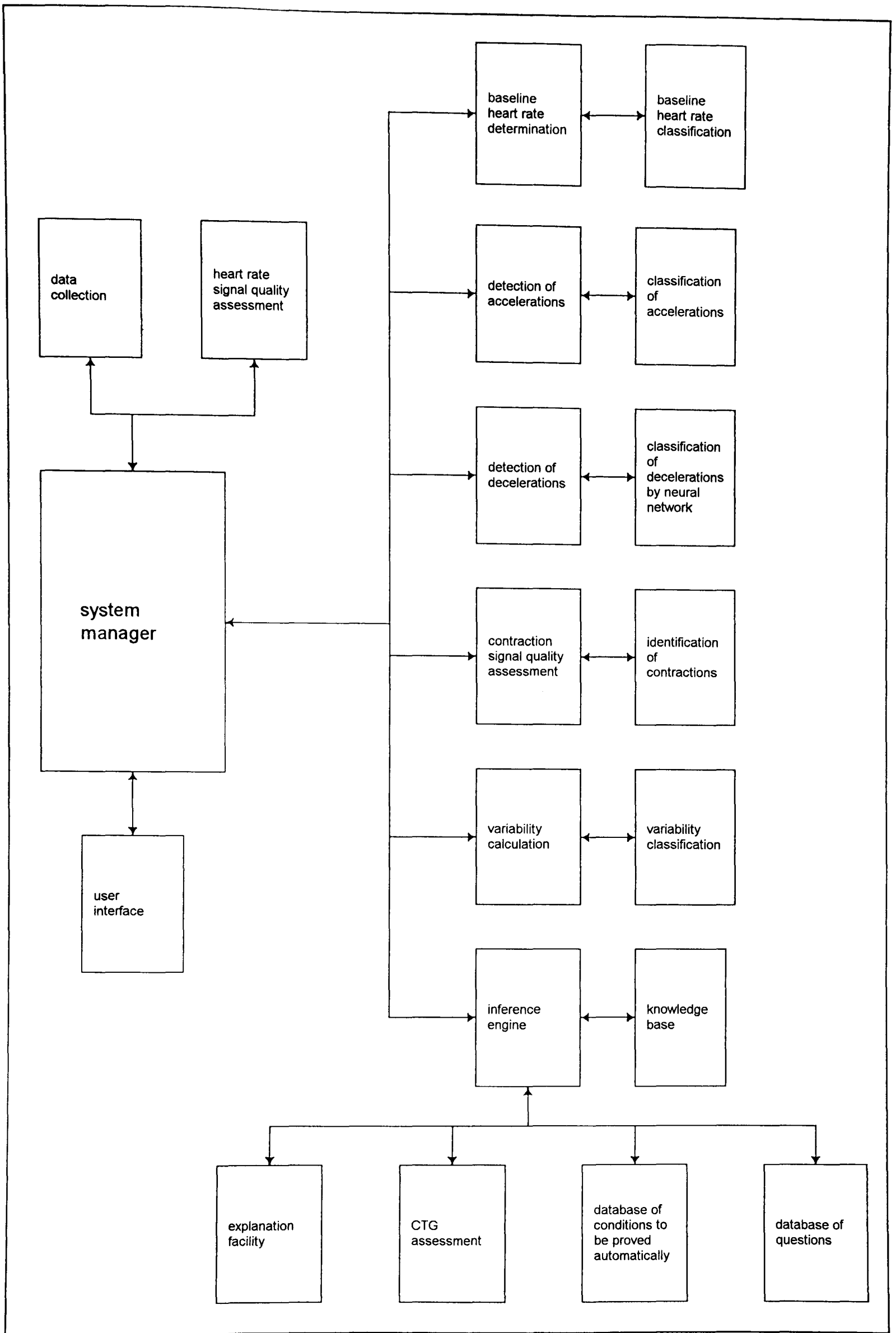


Figure 4.9: System control.

Chapter 4 summary.

Chapter 4 described the development of the expert system module of the intelligent system. Initially, knowledge elicitation took place with two expert obstetricians. From these discussions it was possible to quickly develop a limited expert system. This was used to stimulate the experts to recommend further refinements to the existing knowledge and to also identify areas which required extending. The performance of this prototype system was compared with one of the experts and demonstrated the feasibility of representing expert knowledge for labour management (Ifeachor et al, 1991). This prototype was developed using Prolog software which minimised development time. However, this language was not suitable for the final implementation of the expert system which interfaced to the feature extraction methods. This implementation required the advantages of a low-level language such as 'C' and so the expert system was redesigned. Over a period of two years, the knowledge was refined and represented in tree diagrams and converted into production rules to form the knowledge base. The feature extraction methods described in chapter 3 were integrated with the expert system to obtain the complete intelligent system.

This combined system was a hybrid which utilised the strengths of various techniques used in applied artificial intelligence. The features were extracted from the CTG using numerical algorithms and were classified with additional algorithms and a neural network. These features, together with the patient specific information, formed the dynamic knowledge which was symbolically represented using obstetric terminology (e.g. reduced variability, late decelerations etc.). This dynamic knowledge was then processed by the expert system which applied long term knowledge in the form of the production rules represented in the knowledge base to infer facts and relationships which ultimately lead to actions being recommended.

Chapter 4 references.

- Boose J H (1984) Personal construct theory and the transfer of expertise. *Future Comp Sys.* 1: 2, 143-149.
- Cullen J and Bryman A. (1988) The knowledge acquisition bottle neck: time for reassessment? *Expert Systems* 5: 3, 216-225.
- Diaper D. (1989) Knowledge elicitation: principles, techniques and applications. Ellis Horwood, Chichester, UK.
- Davies M and Hakiel S. (1988) Knowledge harvesting: A practical guide to interviewing. *Expert Systems.* 5: 1, 42-50.
- Hayward S. (1987) How to build knowledge based systems: techniques, tools and case studies. In: ESPRIT '87: Proc of the Esprit conference. Commission of the European Economic Community, Brussels. 665-667.
- Ifeachor E C, Keith R D F, Westgate J and Greene K R. (1991) An expert system to assist in the management of labour. *Expert Systems World Congress Proceedings.* Orlando Florida. Dec 16-19 Liebowitz J. (ed). Pergamon Press, Oxford UK. 4: 2615-2622.
- Keith R D F, Westgate J, Hughes G W, Ifeachor E C and Greene K R. (1993) Preliminary evaluation of an intelligent system for the management of labour. *J Perinat Med.* (Accepted for publication)
- Meyer M and Booker J. (1991) Eliciting and analysing expert judgement, a practical guide. Boose J and Gaines B (eds) Academic Press London UK.
- Neale I M. (1988) First generation expert systems: a review of knowledge acquisition methodologies. *The knowledge Engineering Review* 3: 2, 105-145.
- Seaver D A. (1976) Assessments of group preferences and group uncertainty for decision making. Social Science Research Institute, University of Southern California Los Angeles California.

Chapter 5

Evaluation and validation of the system.

5.1	Introduction	121
5.2	Preliminary evaluation of the expert system	121
5.3	Preliminary evaluation of the integrated system	124
5.4	Validation of the system	128
5.5	Validation results	138
Chapter 5 summary		177
Chapter 5 references		179

5.1 Introduction.

Evaluation and validation are recognised as the most important aspects in the development of intelligent systems (O'Keefe et al, 1987; O'Leary et al, 1990). Objective assessment at the end of each major stage of development is vital for examining the system's limitations, strengths and weaknesses, which provides valuable feedback and helps shape future work. This process was considered especially important for a system applied to the sensitive area of fetal monitoring. The following examinations were proposed to assess the system for labour management at the important stages in its development.

1. Evaluation of the expert system.
2. Internal evaluation of the integrated system.
3. External validation of the system.
4. Internal clinical trial.
5. External randomised trial.

Stages 1, 2 and 3 are described in this thesis. These compare the system with domain experts to determine the accuracy of the embedded knowledge and the validity of the system's advice. The outcome of these investigations would indicate whether stages 4 and 5, which examine the impact of the system in the clinical situation, were viable.

5.2 Preliminary evaluation of the expert system.

5.2.1 Introduction.

A limited expert system was quickly developed after the knowledge elicitation sessions had identified the essential aspects of labour management. This was developed in Prolog software and is described in chapter 4 (section 4.4). The automatic methods for feature extraction from the CTG were not incorporated. Instead, a description of the important features was provided by a midwife. The purpose of this expert system was to provide feedback and define areas for future development. This system was then evaluated by comparing its performance with one of the experts principally involved in its development. The objectives were to establish whether it was feasible to formalise expert knowledge for labour management. This work was presented at an international conference and published in full in the refereed conference proceedings (Ifeachor et al, 1990).

5.2.2 Method.

A group of 31 patient records which had received medical intervention, in the form of at least fetal blood sampling, were chosen randomly from our database. This intervention indicated that at some stage during each labour there existed some concerns for the fetus. Each record was examined retrospectively but blind to outcome by the clinical expert, CE, and a midwife using the expert system, ES.

The CTGs from these cases were assessed in 10 minute segments. As automatic feature extraction had not been incorporated, the midwife was asked to describe 4 features of the CTG to the ES for each 10 minute segment. If required, the ES asked the midwife for further patient specific information and then would recommend a course of action to be taken which the midwife noted. The action recommended by the expert system, the CE and the clinical action taken during labour were subsequently compared.

5.2.3 Results.

The condition of the baby at birth in each case was assessed using the acid base status of the cord artery and vein together with Apgar scores measured at 1, and 5 minutes.

The breakdown of fetal blood sampling carried out or recommended in the 31 cases is given in table 5.1 and the overall results are given in table 5.2.

Number of FBS done per case	Number of cases done clinically	Number of cases recommended by CE	Number of cases recommended by ES
1	18	7	3
2	10	5	6
3	3	4	4
Totals	31	16	13

Table 5.1: Breakdown of fetal blood sampling.

	Number of cases clinically	Number of cases recommended by CE	Number of cases recommended by ES
Number fetal blood samples	31	16	13
Operative delivery	15	13	12
Number of babies with cord artery pH < 7.15	8	2*	2*
Number of babies with Apgar score ≤ 7 at 5 minutes	6	2*	2*

Table 5.2: Summary of results.

* These represent the likely number assuming the recommended actions had been taken clinically.

This study found that;

1. In the 31 cases where at least 1 FBS was obtained clinically, less than half were recommended by the clinical expert (15 cases out of the 31 prevented) and the expert system (18 out of 31 cases prevented). These reductions would not have adversely affected perinatal outcome as the babies involved were delivered with a good outcome.
2. Both the CE and ES did not recommend intervention in the same three cases for which an operative delivery for diagnosed fetal distress was carried out clinically. As the actual outcome in these cases were not consistent with fetal compromise, the recommended actions of the CE and ES were more appropriate.
3. The CE and ES recommended an extra fetal blood sample in the same three cases which would have allowed an earlier detection of fetal distress and an earlier operative delivery by up to 60 minutes. This may have led to an improvement in perinatal outcome.
4. In a further case, an operative delivery was recommended by both the CE and ES but was not performed clinically.
5. An extra FBS was recommended by the CE in 2 cases and by the ES in 3 cases (the same 2 + 1) than was obtained clinically.

From a subsequent review of the cases, it was considered that the small differences between the CE and ES had resulted from differences in the visual interpretation of the CTG.

5.2.4 Discussion.

In this preliminary evaluation of 31 cases where intervention had occurred clinically, it was found that the performance of the prototype expert system was comparable with the clinical expert's. This demonstrated the feasibility of formalising expert knowledge for labour management. Compared with the actions taken clinically, both the clinical expert and expert system specified a lower intervention rate in the form of fetal blood sampling (< 50%) and

operative delivery (20%). In addition, both recommended earlier delivery in the same 4 cases which may have resulted in an improvement in outcome as measured by the cord artery pH and Apgar score at 1 and 5 minutes. However, no fetus in the cases considered had a particularly poor outcome.

5.3 Preliminary evaluation of the integrated system.

5.3.1 Introduction.

This evaluation took place after the expert system had been redesigned and integrated with the automatic feature extraction methods. The purpose of this investigation was to establish whether the combined system could manage labour with a performance comparable with experienced clinicians in Plymouth who had not previously been involved in the development of the system. This study formed the basis of a journal paper which has been accepted for publication (Keith et al, 1993).

5.3.2 Method.

A randomised trial in Plymouth had previously assembled the case notes and CTGs, and measured the outcomes for 2400 labours over a period of 18 months. During the trial, a computer was connected to a single intrapartum recorder at bedside which digitised and recorded the CTGs from 300 labours. One hundred of these cases contributed to the development of the feature extraction methods used by the system. From the remaining 200 cases, the 9 cases which had clinical intervention for 'fetal distress' (fetal blood sampling and/or operative delivery) were selected together with a further 21 cases selected randomly.

The clinical decisions taken in these 30 cases were compared with those recommended in a retrospective review of the cases by the system and 3 experienced clinicians (A, B and C) who were considered to be experts. Experts A and C had not previously been involved in the system's development. The reviewers assessed the cases independently, blind to outcome and with no knowledge of the system's recommendations or each others. The CTGs were exposed in 15 minute segments and each reviewer was asked to make an assessment of the condition of the fetus based on the newly-revealed segment, the previous recording and the clinical information known at the time. They were then asked to specify an appropriate course of management; continue with the labour, obtain a FBS, or request operative delivery. It was not possible for any reviewer to see future segments before the present segment had been assessed. The reviewers were made aware of any additional

information gained from clinical actions (i.e. fetal blood sampling) if they too had specified the same action.

The consistency of reviewer A was also assessed in a second assessment (A2) of the same cases five months later with the case order randomised. Reviewers B and C took part in a discussion of the results of the first assessment and were thus too familiar with the cases to make their subsequent review useful.

5.3.3 Results.

In 18 cases, no action was taken clinically nor was recommended by any reviewer nor the system. For these cases, the mean cord artery pH was 7.26 ($\sigma = 0.07$) and the vein pH was 7.35 ($\sigma = 0.06$). The one minute Apgar score was 8 or more in 16 cases, 6 in one case and 5 in the other. The five minute Apgar scores were 9 or 10 in all cases. These measures indicated that a good outcome had been obtained and that intervention was unnecessary.

The 12 remaining cases where intervention occurred clinically, or was recommended by the system or a reviewer, are shown in table 5.3. The mean cord artery pH was 7.19 ($\sigma = 0.08$) and vein was pH 7.27 ($\sigma = 0.08$). These measures were not significantly different from the non-intervention group, although the trend was towards a lower pH.

Case	Fetal blood samples						Operative delivery						Perinatal outcome					
													Cord blood pH		Base deficit		Apgar	
	Sys	A	A2	B	C	Clin	Sys	A	A2	B	C	Clin	artery	vein	artery	vein	1'	5'
1	2	1	1	2	2	2	cs	cs	cs	cs	cs	cs	7.21	7.31	3.1	7.1	7	10
2	1	1	1	1	0	1	cs	cs	cs	cs	-	cs	-	7.26	-	7.1	7	9
3	1	0	0	0	1	0	-	-	-	-	-	-	7.26	7.29	1.2	2.8	6	8
4	0	0	0	0	1	0	-	-	-	-	-	-	7.13	7.21	8.8	10.0	8	9
5	1	0	0	1	0	1	for	for	for	for	for	-	7.07	7.27	7.4	7.0	9	9
6	0	0	0	0	0	1	-	-	-	-	-	-	7.29	7.36	4.6	4.1	7	9
7	0	0	0	0	0	0	for	for	for	for	for	for	7.23	7.33	1.8	0.9	6	9
8	0	0	0	0	0	0	for	for	for	for	for	-	-	7.12	-	11.2	9	9
9	0	0	0	0	2	2	-	-	-	-	-	-	7.27	7.36	0.9	1.1	8	9
10	1	1	0	1	1	0	cs	cs	cs	cs	cs	cs	7.07	7.12	7.2	5.2	3	8
11	0	0	0	0	0	2	-	-	-	-	-	-	7.20	7.38	3.5	4.0	8	9
12	2	2	1	2	2	2	cs	cs	cs	cs	cs	cs	-	7.29	-	3.9	9	9

Table 5.3: Results from preliminary evaluation of the system.

This table details the interventions and outcomes for the 12 cases, of the 30, where intervention took place clinically or was recommended by a reviewer or the system. Where, sys = system, Clin = clinical action, cs = caesarean section and for = forceps.

1. In no case, of the 30, did the system recommend an action not also recommended by an experienced reviewer.
2. All reviewers and the system recommended an operative delivery in the same three cases with lowest cord pH's (cases 5, 8 and 10); in cases 5 and 8, a forceps delivery was recommended 30 and 50 minutes respectively, before a spontaneous vaginal delivery occurred. In case 10, a FBS was recommended by all reviewers and the system (except A2) 40 minutes before the CTG severely deteriorated and an emergency caesarean section was performed.
3. Operative intervention was recommended in the same 7 cases by the system and all reviewers within ± 1 segment (15 minutes), with one exception. The CTG in this case (case 2), revealed an unstable baseline heart rate from the start of the recording. The system and reviewers A and B, interpreted this recording as a normal baseline with decelerations whilst reviewer C alone, interpreted the trace as a low baseline with accelerations. On the basis of their interpretation, all except C recommended a FBS that was obtained clinically and found to have a pH of 7.15 indicating that immediate delivery by caesarean section was necessary. This case was a good illustration of the importance of correctly determining the baseline heart rate which is essential for the correct classification of periodic changes and therefore fetal condition.

The actions recommended by the system, the reviewers, and the actual clinical management of the 30 cases were further compared. An agreement in management was considered to occur when all recommendations for fetal blood sampling and recommendations for operative delivery were specified within ± 1 segment (15 minutes). The number of case agreements out of 30 formed on this basis, are shown in table 5.4.

	Number of agreements				
	System	A1	B	C	Clinical
System	-	27	29	25	22
A1	27	-	28	25	22
B	29	28	-	24	23
C	25	25	24	-	21
Clinical	22	22	23	21	-

Table 5.4: The reviewers agreement matrix.

The actions recommended by reviewer A in his two reviews (A1 and A2) were compared and found to be entirely consistent in both action and timing in 28 out of 30 cases. The

inconsistencies occurred in cases 10 and 12 and arose from differences in fetal blood sampling. However, for these cases, reviewer A still recommended operative delivery in the same 15 minute segment as in the first review.

5.3.4 Discussion.

The integrated system achieved good agreement with 3 experienced clinicians for reviews of 30 complete labours. Every action recommended by the system was also recommended by at least one experienced reviewer. The highest level of agreement was with reviewer B whose knowledge has been principally incorporated into the system. However, the system also agreed well with the recommended management of reviewers A and C, who although from the same Unit, had not previously been involved in the development of the system.

This limited study also found that the experienced reviewers tended to agree in their management of the labours and expert A was entirely consistent in his recommendations for operative intervention. There was less agreement between the actual clinical actions taken on the labour ward and those recommended by the reviewers.

5.4 Validation of the system.

5.4.1 Introduction.

The possible limitation of a single centre developing a system of this type in isolation was recognised. Therefore, it was decided to invite the assistance of 17 practising experts from 16 leading centres in fetal monitoring within the UK, to assist in a study to validate the system. A letter was sent to the Head of Department of each centre inviting their collaboration. If agreeable, they were asked to nominate an individual regularly involved in the interpretation of the CTG and the management of labour who they regarded as expert.

5.4.2 Objectives.

The objectives for this study were to

1. Compare the recommended management of the system with the 17 experts in a large number of cases to establish if the system could manage labour with a performance comparable with experts.
2. Investigate whether experts could agree on the management of labour.
3. Investigate whether the experts could be consistent in labour management.
4. Establish whether the management of the system was independent of the user's obstetric knowledge.

5.4.3 Method.

The case notes and CTGs were obtained for 50 cases selected from a database of 2400 high risk labours in which the perinatal outcome had been measured by both cord blood gas analysis and Apgar scores. None of the cases had previously been reviewed by the system. The number of cases chosen was considered the maximum that an expert could review in a single day. The cases included a wide range of outcomes, from the birth asphyxiated to the very normal. This study was intended to fully test the ability of the experts and the system and to that end, the 50 cases contained the cases considered to be the most difficult in the database. The minimum length of CTG recording was 2 hours, the maximum was 15 hours and the average was 6 hours. A summary of these cases is given in table 5.5 which details the actual mode of delivery and the measured perinatal outcome.

Case	Mode of Delivery			Perinatal Outcome					
	Vaginal	Forceps	C-section	Cord Artery		Cord Vein		Apgar scores	
				pH	BDecf	pH	BDecf	1'	5'
1		•		7.18	6	7.29	4	9	9
2			•	7.32	1	7.36	2	9	9
3		•		6.93	17	7.12	10	4	9
4		•		7.20	5	7.29	4	5	9
5		•		7.21	4	7.30	3	5	9
6			•	6.96	14	7.00	10	4	9
7	•			6.97	8	7.14	8	2	8
8		•		7.25	4	7.31	6	8	9
9			•	7.03	13	7.10	13	5	7
10			•	7.05	5	7.12	6	3	7
11	•			-	-	7.17	6	9	9
12	•			7.28	3	7.39	3	9	9
13			•	7.28	3	7.32	4	5	9
14	•			7.25	0	7.32	1	5	9
15	•			6.87	16	7.09	13	5	7
16			•	7.05	-	7.13	7	7	9
17			•	-	-	7.21	1	6	9
18			•	-	-	7.35	3	9	9
19		•		7.15	5	7.23	3	9	9
20	•			7.06	3	7.08	5	6	9
21	•			7.25	10	7.42	2	9	9
22	•			7.19	4	7.24	6	5	9
23			•	7.30	0	7.37	0	8	9
24			•	7.17	2	7.22	4	6	9
25	•			6.87	17	6.96	18	6	7
26	•			6.95	7	7.13	4	6	9
27	•			-	-	7.25	5	9	9
28		•		6.97	7	7.10	6	5	9
29			•	7.13	4	7.20	6	6	6
30	•			6.88	22	6.97	15	6	9
31	•			7.25	0	7.37	0	9	9
32			•	7.15	10	7.21	10	6	9
33	•			7.31	0	7.38	2	8	9
34		•		7.14	7	7.20	5	4	7
35			•	6.81	11	7.20	-	3	8
36		•		7.07	4	7.30	1	4	9
37		•		7.19	7	7.22	6	8	9
38	•			7.04	11	7.08	11	3	6
39	•			7.38	8	7.40	6	9	9
40			•	7.25	3	7.27	6	7	9
41	•			6.97	12	7.06	7	9	9
42		•		7.29	-2	7.34	-1	9	9
43			•	7.13	5	7.17	5	7	9
44			•	7.11	8	7.15	7	6	9
45		•		7.22	4	7.36	0	9	9
46	•			7.23	2	7.29	4	9	9
47	•			-	-	7.39	1	9	9
48		•		7.20	4	7.25	3	6	7
49	•			7.35	-1	7.43	3	8	10
50	•			7.32	-1	7.39	0	9	9

Table 5.5: Mode of delivery and perinatal outcome in the 50 cases used for validation.

The objectives for this study meant that its design would need to be carefully conceived to simulate, so far as possible, the clinical situation. Close attention to detail would be required to ensure that the experts had all the relevant clinical information they would need to indicate how they would have managed a given case. It was also important to ensure that this information was presented at the appropriate time during each review. In this way, the experts would have no more and importantly, no less advantage over the clinician actually managing these cases on the labour ward. The information which was considered important included the CTG, a synopsis of each patient's obstetric history, the specific events that occurred during each labour and importantly, an option to obtain further information on the condition of the fetus with a fetal blood sample.

The experts could not review the CTGs from the selected cases directly as they were legal documents and contained annotations made by clinical staff during the labour. To overcome this, each 30 minutes of CTG recording were photocopied and the annotations removed using white paint. These were then re-photocopied, cut out and stuck together to reform a continuous recording. A red line was drawn and numbered on the CTGs to indicate every 15 minute segment which acted as the time reference for each case. With the case notes, all the relevant clinical details were written on the trace within the appropriate time markers. These details represented the clinical information an obstetrician would normally be aware of if actually managing the cases, for example, the administration of drugs and anaesthetics, the cervical dilatation measured during vaginal examinations, the descent and presentation of the baby, the colour of the amniotic liquor etc. The CTGs were then rolled up and stored. A synopsis of each patient's obstetric history was then compiled from the case notes. This synopsis, together with the labour events and perinatal outcome for each case are presented in appendix H.

Two graphs were also constructed for each case. The first plotted the cervical dilatation over the duration of the labour using the data recorded from vaginal examinations. The second graph plotted the estimated fetal blood sample pH for every segment of recording which were obtained as follows. All cases had blood gas analysis at birth so the end point of each graph was fixed. If any blood samples were obtained clinically, then these were plotted and the intermediate points interpolated. If no FBS were taken during the labour, the graph prior to the end point was estimated by 2 obstetricians from a knowledge of the end point and an interpretation of the CTG.

The experts were working in hospitals all over the UK (figure 5.1). The location of the hospitals were split into 4 regions; the South, Scotland, the North-West and North-East. Four workers from Plymouth were given the responsibility of visiting the experts in each area. Two sets of the CTGs and clinical data were compiled which meant that any two workers could be visiting the experts at any one time.



Figure 5.1: The location of the centres.

5.4.4 Study protocol.

Each expert reviewed the 50 cases twice, at least one month apart. The reviews were independent and blind to perinatal outcome and no expert had a knowledge of any other expert's identity. The cases were presented to all experts in the same sequence but the sequence was randomly changed between the two reviews and the cases were reassigned with different identifying numbers.

A method was devised to simulate the clinical situation by revealing the CTGs (with the labour events written in the appropriate segments) from a sealed opaque box in 15 minute segments (figure 5.2). This method prevented an expert seeing future segments before the current segment had been commented upon and also prevented them from estimating the length of the CTG which would have provided a clue to the length of the labour.

The experts scored each segment according to the concern they had for the fetus and the management they considered most appropriate. The scores they could assign were,

Score	Comment
1	I am not concerned for this fetus.
2	I have concerns for this fetus, but they are not sufficient to request a FBS. I may take some remedial action (specify).
3	I am sufficiently concerned to request an FBS or if possible, a simple vaginal delivery.
4	The information I have leads me to be seriously concerned for this fetus. I am not going to recommend immediate delivery although I am thinking delivery and will do so if things deteriorate further.
5	I am so concerned for this fetus that I want immediate delivery.

Scoring sheets were provided to allow the experts to record their score for each segment. In addition, space was provided alongside each entry for the experts to note any comments they wished to make for example, why they had recommended a particular score, what remedial action they would have taken if they had been managing the labour etc. The review of each case was concluded when either the CTG ended or a score was reached which permitted delivery (5 or 3 in the second stage). If an expert requested an FBS in a given segment (score 3) then the referee could provide a result secretly read from the FBS graph. In the same way, if an expert considered a vaginal examination was appropriate to check on progress, then the referee could provide a measure of the cervical dilatation. The aim for the experts, as in the clinical situation, was to achieve minimal intervention without jeopardising the safety of the fetus.

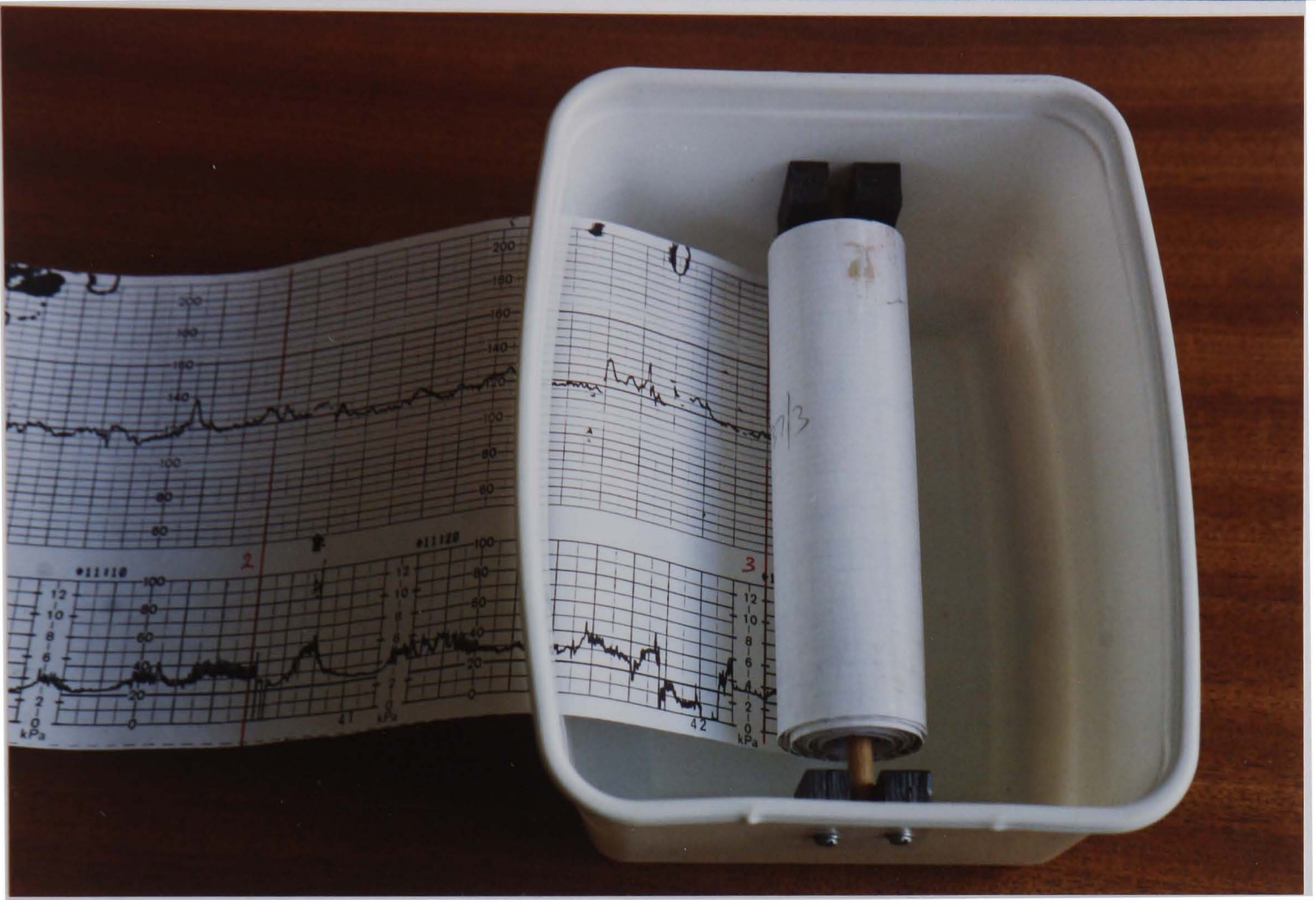


Figure 5.2: The method for revealing the cardiotocograms.

To ensure that each review took place under similar conditions, guidelines were produced for the referees and the reviewers which they read before each review commenced.

Reviewers instructions.

- The CTGs will be revealed to you in 15 minute segments.
- All labour information is contained on the CTG.
- If you wish an FBS or a VE, the referee will provide you with a result on which you may base your subsequent management.
- Score each segment according to how concerned you are for the fetus. A score may carry with it a management procedure.

Score

Comment

- 1 I am not concerned for this fetus.
- 2 I have concerns for this fetus, but they are not sufficient to request an FBS. I may take some remedial action (specify).
- 3 I am sufficiently concerned to request an FBS or if possible, a simple vaginal delivery.
- 4 The information I have leads me to be seriously concerned for this fetus. I am not going to recommend immediate delivery although I am thinking delivery and will do so if things deteriorate further.
- 5 I am so concerned for this fetus that I want immediate delivery.

- You may score each segment as you wish, there are no restrictions.
- Each case review shall end when either the CTG runs out or you reach a score which permits delivery (3 or 5).

Note.

1. We are not particularly interested in decisions to deliver purely for "failure to progress". If you feel you would have delivered for this reason, but are not particularly concerned for the fetus (score 1 or 2), then make a note of this and continue with the case.
- 2 Colour of liquor is stated at the time of SRM or ARM. Unless specifically mentioned, assume no change has occurred in colour during the course of labour.
3. Fetal growth is judged to be normal unless specifically stated otherwise.
4. The point at which an FSE is applied is marked on the CTG, until then the heart rate is derived using ultrasound.

Referees instructions.

- Present the reviewer with the practice case. This is to be treated as a real case although the reviewer should be informed that it is a practice.
- After the practice case, the cases must be presented in the correct order.
- The reviewer is to read each case history prior to each case review.
- The referee must secretly load each CTG into the opaque box to prevent giving an indication of the trace length to the reviewer and must pull out the trace from the box.
- The reviewer shall record their own scores and comments, but the referee must observe this.
- The reviewer may record any comments they wish but do ensure that all FBS and VE results are recorded. Record also the case time at which they asked for it as they may require this information later. Take this time as the "end of segment time".
- Assume an easy vaginal delivery is possible when the woman is 10cm dilated (on graph). At this point, a concern score of 3 represents an easy vaginal delivery and the case review is concluded.
- Each case review is complete either when the CTG runs out or the reviewer reaches a permitted delivery score (3 or 5).
- We want all 51 cases to be reviewed (including the practice case). However, if time runs out we need an absolute minimum of 31 cases .

5.4.5 The system's review.

The CTGs from these cases had not been digitised at the time they were recorded and so a method was required to extract the fetal heart rate and uterine contraction data from the paper CTGs. This was accomplished using a Hewlett-Packard flat-bed colour scanner. Each 30 minutes of CTG recording (of which there was some 300 hours in total) was scanned as a black and white drawing (i.e. no grey scales). This format made use of a threshold filter within the scanner software to obtain an image made up of pixels represented by 0 or 1 (black or white). The threshold was set to ensure the heart rate and contraction tracings were obtained and the background was rejected. Each 30 minute image was viewed and edited to remove all noise and clinical annotations. The 'cleaned' image was then saved in the TIFF image format (tagged image file format, V.5.0). During scanning, a template was used to ensure that the scanning region was precisely to the edges of the CTG paper. With the boundaries of the image known, the relative location of the required data within the image was identified. The heart rate and contraction data samples could now be identified as the black dots within each white scan line. As the correspondence between each scan line and the original CTG were known, it became a simple matter to develop software to extract and scale the required data and save it in a binary file format required by the system. Test

patterns were used to confirm that this method was accurate and every extracted 30 minute epoch was displayed on a computer screen and visually compared to the original CTG.

It was required to directly compare the system's recommended management of the cases with the experts. This was achieved by coding each node in the knowledge tree with the appropriate protocol score, prior to the commencement of the study. The system then displayed this score for each 15 minute segment of recording.

The system reviewed each case twice, with a different operator for each review. The role of the operator was to provide the additional case information and estimates of FBS results when requested by the system. In the first review the operator was an engineer who understood obstetric terminology. In the second review, the operator was an obstetrician not previously involved in the development of the system. The reviews were independent and the operators had no knowledge of each others results.

The scores recorded by the experts and the system for each case were entered in a spreadsheet. All entries were double checked by two workers to ensure they had been accurately entered. The spreadsheet obtained for case 1 is shown in table 5.6 and the remainder can be found in appendix H. The experts were coded, A to Q inclusive and the system was coded, S, with each of the two reviews identified as 1 and 2. Two sets of plausible random numbers were also generated for each case, denoted by R1 and R2. These will be considered in detail later, but their purpose was to indicate the likely results of a reviewer randomly assigning their scores. In addition, the actual clinical actions in the cases were also entered where possible.

It will be recalled from the protocol that a score of 3 represented a request for a FBS if recorded during the first stage of labour and an easy forceps delivery if recorded during the second stage. These actions were scored similarly because a simple forceps delivery and a FBS were considered to be equal interventions. However in reality, a simple forceps delivery cannot always be obtained in which case a FBS may still be preferred before attempting a difficult forceps delivery. The experts could not be told whether a forceps delivery was likely to be easy or difficult because this depended on many factors including the descent and orientation of the baby. These details were not recorded in the clinical notes and so if the reviewers reached a protocol score of 3, it was always assumed a simple forceps delivery was possible. In the spreadsheets, scores of 3 in the second stage were distinguished from FBSs by re-coding them as 6 but as they were considered to be similar interventions, they were left unchanged during the numerical analysis.

5.5 Validation results.

The objectives for this study were to investigate whether an intelligent system could be developed to obtain a comparable performance to experts in the management of labour and whether experts, when given all the relevant information, could largely agree and be consistent. To examine this, it was necessary to investigate the following;

1. Analysis of the recorded scores.

To establish how similarly and consistently the experts and the system managed the 50 labours.

2. Cases where a caesarean section was recommended.

To assess the similarity and consistency of the experts in recommending a CS and whether the system could obtain a similar performance.

3. Fetal blood sampling rates and second stage interventions.

To obtain additional information on the intervention rates of the system and experts.

4. Cases which obtained a poor outcome clinically.

To establish whether the system and experts could have avoided cases which achieved a poor outcome clinically.

5. Cases which obtained a good outcome clinically.

To establish the levels of unnecessary operative intervention specified by the system and the experts.

5.5.1 The distribution of scores recorded by the experts and the system.

The frequency of each reviewer's scores were found and shown in table 5.7. Second stage interventions were re-coded as 6 to separate them from fetal blood samples (score 3). The mean of the experts' frequencies was calculated and are also shown. The experts were labelled A to Q inclusive. The system is labelled S and the plausible random numbers R. The two reviews were denoted by 1 and 2.

Reviewer	Frequency of scores					
	1	2	3	4	5	6
A1	366	535	87	45	20	5
A2	494	535	62	21	17	7
B1	404	322	82	113	27	10
B2	434	346	84	86	25	14
C1	601	292	63	34	18	17
C2	574	379	77	26	17	20
D1	582	347	71	15	20	15
D2	476	455	61	31	18	13
E1	371	515	94	12	21	15
E2	305	606	77	14	19	12
F1	719	269	47	17	17	8
F2	706	265	50	23	18	12
G1	441	600	43	24	15	11
G2	356	592	41	106	16	10
H1	391	408	74	17	25	14
H2	416	488	71	4	24	14
I1	438	573	76	13	16	9
I2	364	640	82	7	16	17
J1	285	613	81	24	18	13
J2	453	486	65	23	19	13
K1	629	283	77	21	17	14
K2	693	241	69	57	17	12
L1	459	487	86	6	18	18
L2	517	432	60	25	18	16
M1	623	501	29	0	10	6
M2	508	570	47	0	12	13
N1	391	542	57	82	16	16
N2	419	526	69	61	16	18
O1	487	390	92	11	20	14
O2	564	422	88	0	16	18
P1	533	489	34	28	17	7
P2	503	539	42	10	16	9
Q1	171	545	0	195	16	12
Q2	54	635	0	279	15	12
S1	685	356	35	30	11	10
S2	674	366	34	29	11	10
R1	462	431	67	41	12	4
R2	495	429	68	30	8	6
\bar{X} (experts)	463	467	63	42	18	13

Table 5.7: Frequency distribution of the recorded scores.

Table 5.7 allowed some preliminary inferences to be made regarding the management recommended by the experts and the system. The average number of fetal blood samples (FBS) per expert review was 63. 7 experts in 9 reviews, requested 80 or more, of which 2 experts (E and O) requested more than 90; an average approaching 2 FBSs per case. In contrast, 5 experts (F, G, M, P and Q) requested 50 or less FBSs in each review; an average of less than one per patient. Of these, expert Q recommended no FBS in either review. In terms of consistency, for 6 experts the difference between the number of FBSs requested in their reviews was less than 5. However, for 2 experts (A and L) the difference in their requests for fetal blood sampling between their reviews was 25 or more. A detailed analysis of FBS rates can be found in section 5.5.4.

The average number of recommendations for immediate CS was 18 per review. 6 experts (A, B, D, E, H and O) recommended 20 or more and of these, 4 experts (A, B, E and O) also requested high numbers of FBS (> 80). Expert M can be identified as a low interventionist. He requested fewer FBSs than most and recommended the least number of CSs.

It is interesting to identify the potential low and high interventionists, but at this stage it is not possible to say which camp have adopted the most appropriate strategy. If the high interventionists consistently identified the compromised fetus early, then there is an argument for their management. However, if the high interventionists performed no better than low interventionists in this respect, then there is an argument against them. The converse is true of the low interventionists. If they had not intervened sufficiently to be able to identify the compromised fetus, then their strategy is inappropriate. On the other hand, if the low interventionists performed as well as the high interventionists in this, then their strategy is preferred.

The argument comes down to accuracy and balance. Accuracy in terms of being able to consistently identify the compromised fetus, and balance in terms of minimising interventions. In this respect, it may be that the remaining experts' strategy is best, that of moderate intervention. This story will unfold.

Expert Q stands out as being quite different from the other experts. He has scored a comparatively low number of 1's, requested no FBSs in either review, and scored a disproportionate number of 4's. In terms of consistency, expert Q was relatively consistent in the number of CSs recommended but was inconsistent in scoring 1's, 2's and 4's. However, these differences in scoring do not tell us whether expert Q performed better or worse or indifferent from the others. This will become apparent with further tests.

The system did not obtain the highest nor the lowest frequency in any score. However its distribution does seem skewed towards the lower scores (less-intervention) compared to the

majority of experts. The plausible random numbers obtained frequencies close to the mean frequencies which confirmed that the numbers were generated with the same probability distribution as the experts' scores.

5.5.2 Analysis of agreement.

A method was derived to measure the agreement between any two reviewer's sequences of scores for a given case. **As this derivation is slightly involved, it is presented in detail in chapter 6.** This measure obtains a value of 100% for perfect agreement and 0% when there is no similarity. The agreement was calculated for all pairs of reviews and were formed in an agreement table for each case. The table obtained for case 1 is shown in table 5.8 and the complete set is given in appendix H. The diagonal, from top left to bottom right represents the calculated agreement of each sequence with itself and confirms that for identical sequences, the agreement measure scores 100%. Table 5.9 can be interpreted as follows. The agreement measured between expert I's first review, I1, and expert A's first review, A1, was 82%. Similarly, the measured agreement between I1 and the system's first review, S1, was 66% and for Q1 and H2 it was 100%.

It was proposed to examine the nature of the agreement results obtained by the experts and the system using two statistical tests.

1. The kappa coefficient of inter-rater agreement.

This measure compares inter-rater agreements and examines the significance of measured agreement in excess of that which could have been expected by chance.

2. The Kruskal-Wallis one-way analysis of variance by ranks.

This non-parametric statistic examines the characteristics of more than 2 distributions to establish whether any differences in the distributions represent genuine population differences or whether the variations are those to be expected from random samples of similar populations. This test will establish if any expert or the system obtained lower levels of agreement than any of the other reviewers for a given case.

The application of these statistics are relatively complex and so are summarised in this chapter but presented in detail in chapter 6 (section 6.5).

Did the experts or system obtain an average agreement in any case significantly better than, or worse than expected by chance?

For each case, the mean agreement the system scored with the experts, $\kappa(\text{system})$, and the mean agreement the experts scored with each other, $\kappa(\text{experts})$, were obtained and are shown in table 5.9. These were expressed as kappa statistics which test the null hypothesis that neither the experts, nor the system, obtained agreements significantly different from those expected by chance. Two alternative hypothesis were considered.

1. If the agreement reached significance and was less than expected by chance then it was concluded that there was significant disagreement.
2. If the agreement reached significance and was greater than expected by chance then it concluded that there was significant agreement.

Although two alternative hypothesis have been formed, the test was one-tailed because each predicts the direction of the agreement. In this test, the region of rejection should not be too small to prevent a measure of agreement from reaching significance (in either direction) when the number segments, n , scored in a case was small. Therefore the significance level was chosen as $\alpha = 0.1$. **The detailed calculations for this test are given in chapter 6 (section 6.5.1).**

Case	Segments (n)	Measured agreement	
		o (experts)	o (system)
1	12	68.1	71.5
2	45	43.7	44.7
3	47	52.2	54.8
4	20	57.9	57.4
5	26	58.7	63.4
6	10	92.1	91.6
7	28	67.8	76.6
8	16	76.5	63.6
9	7	80.5	76.2
10	9	64.4	70.2
11	16	80.6	65.2
12	19	55.7	62.4
13	56	59.4	65.1
14	16	65.0	61.7
15	20	66.3	67.9
16	47	44.0	52.1
17	20	57.7	46.7
18	36	66.6	66.6
19	15	59.1	57.2
20	23	75.0	81.2
21	16	90.6	92.0
22	20	66.3	67.9
23	52	70.0	73.3
24	11	64.0	52.4
25	14	80.4	76.8
26	31	75.9	76.9
27	8	79.1	75.9
28	26	69.5	73.0
29	52	84.5	79.0
30	11	63.9	49.4
31	16	97.1	97.6
32	17	49.8	40.2
33	15	76.2	75.3
34	23	74.9	78.7
35	9	81.5	72.8
36	38	63.7	64.9
37	59	65.8	63.4
38	14	81.3	29.1
39	27	90.8	91.9
40	11	57.7	38.7
41	15	63.6	61.7
42	17	61.7	54.5
43	19	83.1	76.2
44	20	64.8	31.8
45	14	68.7	73.0
46	34	78.4	81.8
47	20	92.5	92.4
48	38	65.0	46.0
49	29	89.2	92.3
50	28	88.0	91.3

Table 5.9: Statistical significance in the measures of agreement for each case.

The average agreement the system scored with the experts, $o(\text{system})$, and the average agreement the experts scored with each other, $o(\text{experts})$, was calculated for each case. These were expressed as kappa values, which were referred to the standard normal tables to test the significance agreement above chance.

it was found that,

1. The average agreement the experts scored with each other was significantly better than was expected by chance in 31 cases.
2. The average agreement the system scored with the experts was significantly better than was expected by chance in 29 cases.
3. The experts significantly disagreed in the management of 1 case (case 2).
4. The system disagreed with the experts in the management of 2 cases (38 and 44).

Did each expert and the system manage each case similarly?

Unlike the previous test which considered average agreements, this test considered the agreement measures obtained by the experts and the system on an individual basis. If an expert had managed a case differently from the other experts, then they would have obtained lower measures of agreement. If the system did not embody expertise then it would be expected to obtain lower measures of agreement with the experts than the experts obtained with each other.

The Kruskal-Wallis analysis of variance test was used to establish if the agreements each reviewer obtained for a given case were similar to other reviewers. **This test is given in detail in chapter 6 (section 6.5.2).** For each reviewer, the number of cases (and reviews) in which the majority of other experts obtained significantly higher agreement was found and is represented in table 5.10. The level of significance for this test was $\alpha = 0.05$.

Reviewer	Number Cases	Number Reviews
A	6	7
B	16	22
C	6	7
D	3	3
E	1	1
F	5	6
G	2	2
H	9	13
I	1	1
J	11	11
K	2	2
L	3	3
M	7	9
N	4	4
O	4	4
P	5	6
Q	28	36
System	5	10

Table 5.10: The number of cases and reviews where each reviewer obtained significantly lower agreements than the majority.

It was found that;

1. The system obtained significantly lower agreement than the majority of experts in 5 cases of the 50.
2. Most experts obtained significantly lower agreement than the majority of reviewers in 5 cases or less.
3. Experts B, H, J and Q, obtained significantly lower agreements than the majority of reviewers most often.
4. Expert Q deserves special mention here because in over half the cases, he obtained significantly lower agreements than the majority.

This test identified those who have managed cases differently from the majority but by itself it does not separate the good performers from the poor performers. It may be that for some cases, those with lower agreements have managed the labours better than the rest.

The results obtained for the system are encouraging because in 45 out of 50 cases, the level of agreement that the system achieved with the experts was not significantly lower than the experts achieved with each other.

Overall, did each expert and the system agree well with the other experts and were they consistent?

Each reviewers average agreement with the other experts (inter-agreement) and consistency (intra-agreement) was calculated. These measures were used to determine whether each expert and the system was able to agree and be consistent in excess of that expected by chance.

The agreement the system obtained with the experts for each of its reviews, S1 and S2, was averaged over all 50 cases. Similarly for each expert review, the average agreement with the remaining experts was found. These are shown in table 5.11 in the column \bar{O}_{inter} . The average of all inter-expert agreements was obtained to give a single measure of the overall inter-expert agreement, \bar{Ex} and the agreement that the two sets of plausible random numbers obtained with the experts was found. The average consistency of each reviewer and the plausible random numbers was taken as the agreement measured between their 2 reviews of each case, averaged over the 50 cases. This is shown in table 5.11 under the column \bar{O}_{intra} .

These results were then expressed as kappa values and used to test the null hypothesis that no expert, nor the system, obtained an average agreement significantly different from that expected by chance. Two alternative hypothesis were considered;

1. If a reviewer's average agreement reached significance and was less than chance then it was concluded that the reviewer had significantly disagreed with the other experts in the management of the 50 cases.
2. If a reviewer's average agreement reached significance and was greater than expected by chance then it concluded that the reviewer had significantly agreed with the management of the other experts in the 50 cases.

The total number of segments, n, the experts scored was 1129 and the level of significance was chosen to be, $\alpha = 0.01$. The analysis of this test is given in chapter 6 (section 6.5.3).

Reviewer	\bar{O}_{inter}	\bar{O}_{intra}
A1	69.61	76.42
A2	73.06	
B1	60.67	78.68
B2	65.18	
C1	71.94	80.00
C2	72.71	
D1	73.13	82.32
D2	73.61	
E1	73.76	82.10
E2	73.35	
F1	66.87	81.20
F2	70.92	
G1	73.40	77.28
G2	70.66	
H1	64.23	78.62
H2	69.76	
I1	73.85	89.04
I2	74.69	
J1	68.93	73.18
J2	72.32	
K1	73.60	81.00
K2	72.04	
L1	73.92	83.16
L2	73.37	
M1	66.02	84.18
M2	70.05	
N1	70.94	82.56
N2	71.48	
O1	69.32	78.44
O2	74.10	
P1	70.01	81.58
P2	72.60	
Q1	59.54	74.16
Q2	56.80	
S1	67.14	99.16
S2	67.51	
\bar{Ex}	70.19	80.23
\bar{Ran}	47.1	50.24

Table 5.11: Each reviewers average inter- and intra-agreement.

The analysis of these results found;

Inter-agreement results.

1. All experts and the system obtained an average agreement significantly better than expected by chance. For the most part (with the exception of Q1 and Q2) these results were highly significant ($p < 0.0000003$).

2. The system's average inter-agreement was within the range of the experts' agreement.
3. Expert Q obtained much the lowest inter-agreement. For his second review, this was only just significantly above that expected by chance.
4. The plausible random numbers significantly disagreed with the experts.

Intra-agreement.

5. The system and the experts obtained a consistency significantly better than could have been expected by chance. Again, these results were highly significant.
6. The system achieved almost perfect consistency (99.16%).
7. The average expert intra-agreement was 80.23% but of these, expert I stood out as being the most consistent (89.04%) and experts Q and J were the least consistent (74.16% and 73.18% respectively).
8. The intra-agreement of the plausible random numbers was not significantly different from that expected by chance (50.24%).

5.5.3 Analysis of decisions to obtain delivery by caesarean section.

There were 31 cases in which at least one expert or the system recommended immediate operative intervention by caesarean section (CS). It was important to examine these cases in detail to establish whether,

1. The system recommended CS when necessary.
2. The system recommended CS unnecessarily.
3. The timing of the system's recommendations was comparable with the experts.
4. The experts agreed (inter and intra) on the cases which should have CS delivery.
5. The experts agreed (inter and intra) on the timing of interventions.

Table 5.12 summarises the actions taken in the 31 cases where at least one reviewer recommended delivery by CS. Each main column heading may be interpreted as follows;

Expert assessment,

- The number of experts recommending delivery by CS in the specific case (max. = 17).
- The number of expert reviews recommending delivery by CS (max. = 34).
- The average segment number ($\overline{segment}$) in which the experts' decisions were taken.

System assessment,

- The operative intervention, if any, the system recommended.
- The segment number in which the system recommended an operative intervention.

Clinical outcome,

- The cord arterial blood gas analysis at birth which includes the pH and Base Deficit of the extra-cellular fluid (BDecf). The babies here with poorest outcomes are those with significant metabolic acidosis (pH < 7.05 and BDecf \geq 12). A detailed assessment of the cases related to outcome can be found in 5.5.6.

Case	Expert Assessment			System Assessment		Clinical Outcome	
	Number of Experts	Number of Reviews	$\overline{\text{segment}}$	Action	Segment	Cord artery	
						pH	BDecf
2	12	18	23.9	none	-	7.32	1
3	15	26	33.5	C/S	41	6.93	17
4	3	4	11.8	none	-	7.20	5
5	7	10	20.6	forceps	26	7.21	4
6	17	33	8.1	C/S	8	6.96	14
7	7	9	25.4	none	-	6.97	8
9	17	34	4.6	C/S	6	7.03	13
10	17	32	4.2	C/S	3	7.05	5
12	4	5	6.2	none	-	7.28	3
13	9	15	51.1	none	-	7.28	3
14	1	1	7	none	-	7.25	0
15	2	3	12.0	none	-	6.87	16
16	17	33	35.2	none	-	7.05	-
17	17	34	12.3	C/S	14	-	-
18	3	4	26.5	none	-	-	-
19	17	34	9.5	C/S	11	7.15	5
22	2	3	12.0	none	-	7.19	4
23	1	2	17.0	none	-	7.30	0
24	16	30	9.4	none	-	7.17	2
28	1	1	3.0	forceps	25	6.97	7
29	17	34	50.4	C/S	51	7.13	4
32	14	26	14.1	none	-	7.15	10
34	5	6	18.7	forceps	20	7.14	7
35	17	34	7.1	C/S	8	6.81	11
37	16	26	46.0	forceps	57	7.19	7
38	17	31	13.1	none	-	7.04	11
40	17	31	8.7	C/S	5	7.25	3
41	3	4	9.8	none	-	6.97	12
43	16	32	15.7	C/S	14	7.13	5
44	17	33	16.3	none	-	7.11	8
48	16	27	33.2	C/S	29	7.20	4

Table 5.12: Summary of cases recommended for caesarean section.

Assessment of each reviewer's recommendations for CS on a case by case basis.

The cases where each reviewer, in at least one of their reviews, recommended immediate delivery by CS were identified. A table for each reviewer was constructed which detailed the intra- and inter- agreement their decisions had received. These tables obtained for expert I, Q and the system are shown in tables, 5.13, 5.14 and 5.15 respectively. The tables for all reviewers are presented in appendix J.

A Guide to interpreting the tables. Consider the table compiled for expert I on the following page. The major headings are,

Case number. Lists all the case numbers where expert I recommended delivery by CS in at least one review.

Intra-agreement. Identifies,

1. The cases where expert I, recommended a CS in both reviews.
2. The relative timing in segments, δ_t , between the decisions in the two reviews.

For example, in case number 10, expert I, recommended a CS in both reviews and the relative timing between these recommendations was 2 segments (30 minutes).

Inter-agreements.

Overall agreements identifies,

1. The number of other experts who recommended a CS in the same case.
2. The number of expert reviews recommending a CS in the same case.

For example, in case 19, 13 other experts in 24 reviews also recommended a CS.

Timing of agreements.

1. This heading is split into two. The first half relates to the CS recommended in expert I's first review, I1, and the second half relates to the CS recommended in expert I's second review, I2. An empty set indicates that a CS was not recommended in that particular review. Each half considers the following;

- (a) How many of the other experts' recommendations for CS fell within ± 1 segment of expert I's recommendation, how many were ± 2 segments away and how many were within ± 4 segments but were further than ± 2 segments.
- (b) If the system too had recommended a CS in the given case, the time, δ_t System, gives the relative difference in timing between expert I's recommendation and the system's.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in I1			δ_t System (segs)	Agreement with C/S specified in I2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
3	YES	0	14	24	11	4	3	2	11	4	3	2
6	YES	0	16	31	31	0	0	0	31	0	0	0
9	YES	0	16	32	23	9	0	1	23	9	0	1
10	YES	2	16	30	19	3	7	0	15	10	5	2
16	YES	0	16	31	5	0	3	-	5	0	3	-
17	YES	1	16	32	14	5	8	2	17	0	5	1
19	YES	0	13	24	11	4	9	2	11	4	9	2
24	YES	0	15	28	17	5	3	-	17	5	3	-
29	YES	0	16	32	30	2	0	0	30	2	0	0
32	NO	-	13	25	-	-	-	-	8	8	8	-
35	YES	0	16	32	29	3	0	2	29	3	0	2
37	NO	-	15	25	3	15	0	-	-	-	-	-
38	YES	0	16	29	24	5	0	-	24	5	0	-
40	YES	1	16	29	6	6	17	2	7	9	13	3
43	YES	1	15	30	25	4	1	2	23	4	3	1
44	YES	2	16	31	14	12	4	-	14	8	7	-
48	YES	0	15	25	6	1	15	9	6	1	15	9

Table 5.13: Detailed assessment of the cases recommended for CS by expert I.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in Q1			δ_t System (segs)	Agreement with C/S specified in Q2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
2	YES	6	11	16	1	2	2	-	0	2	2	-
3	NO	-	14	25	1	1	0	36	-	-	-	-
5	YES	12	6	8	0	0	0	-	2	1	1	-
6	YES	0	16	31	31	0	0	0	31	0	0	0
9	YES	1	16	32	26	5	1	2	23	9	0	1
10	YES	3	16	30	19	4	6	0	10	7	13	3
12	NO	-	3	4	1	0	2	-	-	-	-	-
13	YES	3	8	13	0	0	3	-	1	0	0	-
14	NO	-	0	0	0	0	0	-	-	-	-	-
16	YES	1	16	31	1	1	0	-	1	1	0	-
17	YES	2	16	32	6	3	2	7	9	1	17	5
18	NO	-	2	3	0	0	2	-	-	-	-	-
23	YES	8	0	0	0	0	0	-	0	0	0	-
29	YES	1	16	32	30	2	0	1	30	2	0	0
35	YES	0	16	32	25	5	2	1	25	5	2	1
37	NO	-	15	25	-	-	-	-	4	1	0	-
38	NO	-	16	30	-	-	-	-	5	12	13	-
40	NO	-	16	30	-	-	-	-	7	4	10	1
44	NO	-	16	32	14	9	7	-	-	-	-	-
48	NO	-	15	26	-	-	-	-	15	1	8	5

Table 5.14: Detailed assessment of the cases recommended for CS by expert Q.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS							
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS					
			Num. other experts recommending C/S max = 17	Num. reviews recommending C/S. Max = 34	Agreement for C/S of S1			Agreement for C/S of S2		
					Num. other reviews with C/S within (segs)			Num. other reviews with C/S within (segs)		
± 1	± 2	± 4	± 1	± 2	± 4					
3	YES	0	15	26	9	5	4	9	5	4
6	YES	0	17	33	33	0	0	33	0	0
9	YES	0	17	34	21	5	8	21	5	8
10	YES	0	17	32	20	4	7	20	4	7
17	YES	2	17	34	11	8	3	16	5	8
19	YES	0	14	26	10	6	9	10	6	9
29	YES	0	17	34	32	2	0	32	2	0
35	YES	1	17	34	11	16	7	27	5	2
40	YES	0	17	31	7	3	6	7	3	6
43	YES	0	16	32	18	7	6	18	7	6
48	YES	0	16	27	0	0	1	0	0	1

Table 5.15: Detailed assessment of the cases recommended for CS by the system.

Summary of the assessment of each reviewers recommendations for CS.

1. The system recommended a CS delivery in the same 11 cases in both reviews.
2. The median number of cases that the experts recommended for CS in the first review was 18. The minimum was 10 (expert M) and the maximum was 26 (expert B).
3. The median number of cases that the experts recommended for by CS in the second review was 17. The minimum was 12 (expert M) and the maximum number was 24 (experts B and H).
4. The median number of cases in which the experts were consistent in recommending CS in both their reviews was 15 and the modal value was 15. The range was between 9 and 24.
5. Overall, four experts recommended CS in the same number of cases in both their reviews; four recommended more cases in their second review; 9 recommended less cases in their second review.
6. The system recommended the second lowest number of CS.

Inter- and intra-agreement in decisions to deliver by CS.

Inter-agreement

If a reviewer recommended a CS in a given case, a measure of the agreement for this decision could be represented by the total number of other expert reviews that also recommended delivery by CS. This figure was obtained for each reviewer and then normalised by the total number of other expert reviews which could have recommended this decision. For the experts this total was 32, and for the system and plausible random numbers, the total was 34. This measure of agreement for a reviewer's decision to deliver by CS in a given case is expressed in equation 5.1.

$$CS \text{ agreement} = \frac{\text{Number of reviews recommending delivery by CS}}{\text{Total number of reviews}} \times 100\% \quad (5.1)$$

This figure was averaged over all the cases in which a given expert (or the system) recommended a CS. These figures are shown for each reviewer in table 5.16.

Intra-agreement.

An assessment of each reviewer's intra-agreement was made by comparing the total number of distinct cases where a CS was recommended in at least one review, with the number of cases where a CS was consistently recommended. This measure is represented in equation 5.2.

$$\text{Intra-agreement} = \frac{\text{number cases where CS was consistently recommended}}{\text{total number of distinct cases where a CS was recommended}} \times 100\% \quad (5.2)$$

These figures are also shown in table 5.16.

Reviewer	INTRA -			INTER-
	Total Number of distinct cases with decisions for CS	Number cases consistent between reviews	Intra-agreement (%)	Average inter-agreement (%)
A	21	15	71	76
B	27	23	85	63
C	20	15	75	81
D	20	18	90	82
E	22	18	82	77
F	19	16	84	82
G	19	12	63	83
H	25	24	96	67
I	17	15	88	90
J	21	16	76	78
K	19	15	79	85
L	19	17	89	84
M	13	9	69	95
N	17	15	88	87
O	20	16	80	79
P	18	15	83	90
Q	20	11	55	68
System	11	11	100	92
Random No.	19	1	5	32

Table 5.16: Intra- and inter agreement in decisions to deliver by CS.

Inter- and intra-agreement in the timing of decisions to deliver by CS.

It was also important to compare the relative timings of the decisions to deliver by CS to establish whether the system recognised fetal compromise at the same time as the experts. A comparison of the experts' decision times would help identify those who were consistent.

Inter agreement.

In each case where a given reviewer recommended delivery by CS, the total number of other expert reviews deciding the same action within ± 1 segment, ± 2 segments and ± 4

segments of the given review's decision was found. This total was then normalised by dividing by the total number of reviews recommending CS in the given case. This measure is given by equation 5.3.

$$\text{Inter-timing agreement} = \frac{\text{number of reviews recommending CS within the given time frame}}{\text{total number of reviews recommending CS}} \times 100\% \quad (5.3)$$

An average value for each reviewer over all the cases where they recommended delivery by CS was found and is shown in table 5.17.

Interpretation of table 5.17.

Take reviewer C for example. In all the cases where she recommended delivery by CS, an average of 59% of all other expert reviews also recommending the same action, did so within ± 1 segment (15 minutes) of her decision; 74% did so within ± 2 segments; and 86% did so within ± 4 segments (1 hour).

Expert	Average number of reviews recommending a CS within ± 1 seg (%)	Average number of reviews recommending a CS within ± 2 segs (%)	Average number of reviews recommending a CS within ± 4 segs (%)
A	52	69	86
B	42	61	78
C	59	74	86
D	49	67	84
E	51	67	84
F	55	66	80
G	58	71	85
H	37	52	74
I	58	74	90
J	56	68	84
K	55	72	84
L	61	74	87
M	51	66	82
N	53	67	84
O	53	66	81
P	55	70	83
Q	46	57	71
System	53	67	82
Random No.	13	23	38

Table 5.17: Inter-agreement in the timing of CS.

Intra-agreement

For each reviewer, the average time difference in segments, $\bar{\delta}_t$, was calculated between their CS decision points in the cases where they consistently recommended CS. The standard deviation, σ_t , in timings was also found. This data is represented in table 5.18.

Interpretation of table 5.18.

Consider expert D for example. Expert D consistently recommended delivery by CS in 18 cases. The average difference $\bar{\delta}_t$, between the timings in her two reviews was 2.8 segments and the standard deviation was 3.76 segments,

Expert	Number of cases	$\bar{\delta}_t$	σ_t
A	15	1.5	1.31
B	23	2.3	3.34
C	15	1.7	2.29
D	18	2.8	3.76
E	18	2.0	2.77
F	16	4.6	8.40
G	12	1.6	2.06
H	24	3.3	4.67
I	15	0.5	0.72
J	16	3.3	6.60
K	15	3.9	8.89
L	17	2.6	6.43
M	9	1.1	1.37
N	15	0.9	0.72
O	16	4.0	7.47
P	15	0.8	0.83
Q	11	3.3	3.62
System	11	0.3	0.62
Random No.	1	24	-

Table 5.18: Intra-agreement in the timing of CS.

Summary of results in cases where the decision to deliver by CS was taken.

The system.

1. The system consistently recommended delivery by CS in 11 cases. This was within the range of the experts, but was lower than all but 2 experts (M and Q).
2. On average, 92% of all expert reviews also recommended delivery by CS in these cases.
3. The majority of expert decisions were within 1 segment (15 minutes) of the system's decision and two thirds were within 2 segments (30 minutes). This was comparable with most experts.

4. The system was highly consistent in the timing of CS in its two reviews ($\bar{\delta}_t = 0.3$ segments).
5. There were 7 cases where the majority of expert reviews recommended CS but the system recommended either no action or later delivery by forceps (cases 2, 16, 24, 32, 37, 38 and 44). It was considered important to establish whether the system's decisions here were reasonable. A discussion of these cases can be found in 5.5.8.

The experts.

6. For the cases recommended for CS by 14 of the 17 experts, > 75% of all other expert reviews also recommended the same action. Experts B, H and Q obtained less agreement than most (63%, 67% and 68%, respectively).
7. Expert H consistently recommended delivery by CS in 24 of 25 cases (96%) in both reviews and was the most consistent expert in this respect. Three experts were more inconsistent than most (A, G and Q). Of these, expert Q was consistent in only 11 out of 20 cases (55%).
8. Most experts agreed closely in the timing of CS. However, experts B, H and Q obtained lower agreement for the timing of their decisions.
9. From an examination of the cases where each expert consistently recommended CS in both their reviews, it was found that 11 experts obtained an average difference in timing between their reviews of less than ± 3 segments. Of these, 7 were less than ± 2 segments and remarkably, 3 experts (I, N and P) obtained an average timing difference of less than 1 segment.

The plausible random numbers.

10. As expected, the random numbers obtained little inter- or intra- agreement in either the cases recommended for CS or in the timing of decisions.

5.5.4 Fetal blood sampling rates.

For each review, the total number of patients recommended for an FBS and the total number of FBS requests made was found. This was used to examine whether,

1. The system had FBS rates comparable with experts.
2. All experts had similar FBS rates.
3. The experts were consistent in the cases they recommended for FBS.

This data is summarised in table 5.19.

Review 1			Review 2		
Reviewer	Num. FBS	Num. Cases	Reviewer	Num. FBS	Num. Cases
A1	83	33	A2	60	29
B1	82	37	B2	83	38
C1	62	28	C2	77	34
D1	71	34	D2	61	27
E1	94	36	E2	77	30
F1	47	26	F2	50	27
G1	43	25	G2	41	25
H1	74	36	H2	71	34
I1	76	35	I2	82	37
J1	81	42	J2	65	35
K1	77	33	K2	69	29
L1	86	37	L2	60	29
M1	29	18	M2	47	23
N1	57	28	N2	69	34
O1	92	40	O2	88	38
P1	34	20	P2	42	28
Q1	0	0	Q2	0	0
System 1	35	18	System 2	34	17
Random 1	67	33	Random 2	68	32

Table 5.19: Fetal blood sampling rates for each reviewer.

The consistency of each reviewer's fetal blood sampling decisions was investigated. The total number of inconsistent cases a reviewer recommended for an FBS can be given by equation 5.4.

$$\text{number cases inconsistent} = (n_t - n_1) + (n_t - n_2) \quad (5.4)$$

Where, n_t is the number of distinct cases in which a FBS was recommended in either review, n_1 is the number of cases recommended in the first review and n_2 is the number recommended in the second review. This figure can be interpreted as the number of cases recommended for FBS in the first review but not the second + the number of cases recommended for FBS in the second review but not the first. The results are shown in table 5.20.

Reviewer	n_1	n_2	n_t	Inconsistencies
A	33	29	33	4
B	37	38	40	5
C	28	34	37	12
D	34	27	34	7
E	36	30	37	8
F	26	27	31	9
G	25	25	29	8
H	36	34	38	6
I	35	37	40	8
J	42	35	44	11
K	33	29	34	6
L	37	29	38	10
M	18	23	25	9
N	28	34	36	10
O	40	38	43	6
P	20	28	30	12
Q	0	0	0	0
System	18	17	18	1
Random	33	32	41	17

Table 5.20: The consistency of each reviewer in requesting FBS.

Summary of results in cases where the decision to obtain FBS was taken.

The system.

1. The system recommended 35 FBSs in review 1 and 34 in review 2 from 18 and 17 patients respectively. Only one expert requested a fewer number of FBSs (expert Q).
2. The case in which the system was inconsistent is discussed in 5.5.8.

The experts.

3. Expert Q was significantly different from the other experts in deliberately not requesting a single FBS in either review.
4. There was a wide variation in the numbers of patients the experts considered required a FBS in any one review. Apart from expert Q, the lowest was 18 of the 50 patients (expert M1) and the highest was 40 (expert O1).
5. There was also a wide variation in the numbers of FBS the experts requested in any one review. Apart from expert Q, the minimum was 29 (expert M1) and the maximum was 94 (expert E2).
6. The experts were also relatively inconsistent between their two reviews in specifying the cases they considered required an FBS.

The plausible random number generator

7. As expected, the plausible random numbers were consistent in the numbers of FBS requested (67 and 68) in each review, but were inconsistent in specifying the cases to obtain the FBSs from.

5.5.5 Interventions in the second stage of labour.

It was decided to investigate the reviewers second stage interventions to establish whether;

1. The system's intervention rate was comparable with that of the experts.
2. The experts had similar intervention rates.
3. The experts and the system were consistent in the cases they recommended for intervention in the second stage of labour.

A similar analysis to that for FBS rates was undertaken. The total number of second stage interventions was found for each reviewer in each of their reviews. This data is shown in table 5.21.

Review 1		Review 2	
Reviewer	Num. Interventions	Reviewer	Num. Interventions
A1	4	A2	7
B1	10	B2	14
C1	16	C2	20
D1	15	D2	13
E1	15	E2	12
F1	8	F2	12
G1	11	G2	10
H1	14	H2	14
I1	9	I2	17
J1	13	J2	13
K1	14	K2	12
L1	18	L2	16
M1	6	M2	13
N1	16	N2	18
O1	14	O2	18
P1	7	P2	9
Q1	12	Q2	12
System 1	10	System 2	10
Random 1	4	Random 2	6

Table 5.21: Interventions in the second stage of labour.

A measure of the inconsistencies for a given reviewer was obtained in a similar way to that for the FBS results, using equation 5.5.

$$\text{number cases inconsistent} = (n_t - n_1) + (n_t - n_2) \quad (5.5)$$

Where here, n_t is the total number of distinct cases where second stage intervention was recommended, n_1 is total recommended in review 1 and n_2 is the total in review 2. This data is presented in table 5.22 for each reviewer.

Reviewer	n_1	n_2	n_t	Inconsistencies
A	4	7	11	11
B	10	14	15	6
C	16	20	21	6
D	15	13	15	2
E	15	12	15	3
F	8	12	13	6
G	11	10	13	5
H	14	14	15	2
I	9	17	17	8
J	13	13	17	8
K	14	12	15	4
L	18	16	20	6
M	6	13	14	9
N	16	18	19	4
O	14	18	18	4
P	7	9	12	8
Q	12	12	15	6
System	10	10	10	0
Random	4	6	10	10

Table 5.22: The inconsistencies of each reviewer second stage interventions.

Summary of second stage interventions.

The system.

1. The system consistently recommended 10 cases for intervention during the second stage of labour. This was well within the range of the experts.

The experts.

2. Experts, D and H were the most consistent experts with 2 inconsistencies between the recommendations of their first and second reviews.
3. Expert A was the most inconsistent. In no case did he recommend intervention in the second stage in both reviews.

The plausible random number generator

4. As expected, the plausible random numbers requested a similar number of interventions in each review (4 and 6) but was entirely inconsistent in the cases it recommended.

5.5.6 Intervention in cases with poor perinatal outcome.

The cases discussed here are those where there was evidence to suggest a degree of fetal compromise at delivery. An investigation of these cases will determine whether;

1. The system managed these cases appropriately.
2. The experts managed these cases appropriately.
3. The system and experts performed similarly.

The point was made in chapter 1 (section 1.3.2) that there is no precise measure of outcome. At present, perinatal outcome is assessed using a combination of 3 variables;

- Apgar scores; a subjective score from 0 to 10 where 0 indicates no vital life signs, calculated by clinical staff at 1, 5 and sometimes 10 minutes after delivery.
- Cord blood gas analysis; the pH and base deficit in the extracellular fluid (BDecf) of a sample of blood taken from the cord artery immediately after birth.

Three graded categories of poor outcome will be considered beginning with the most serious.

1. Birth asphyxia.

The most severe category for which there is substantial evidence of fetal compromise. This group was characterised by cord arterial pH < 7.05 and BDecf \geq 12 and Apgar score at 5 minutes \leq 7. There were 3 cases in the study and intervention in these cases can be considered necessary.

2. Severe metabolic acidosis.

These cases had a severe metabolic acidosis but had an acceptable 5 minute Apgar score. This group was characterised by cord arterial pH < 7.05 and BDecf \geq 12 and Apgar score at 5 minutes > 7 (4 cases).

3. Acidosis.

These cases have acidosis but without a significant metabolic component. They are characterised by cord pH < 7.05 but BDecf < 12 (5 cases).

These cases are summarised in table 5.23 and are identified by their case number. The actions taken by the 17 experts are described for each review separately. The system managed these cases consistently and its recommended actions are described.

	Case number	Review 1			Review 2			System's recommendation
		No. of experts recommending			No. of experts recommending			
		CS	Forceps	No Action	CS	Forceps	No Action	
Birth asphyxia	9	17	0	0	17	0	0	C/S
	15	2	1	14	1	1	15	No Action
	25	0	17	0	0	15	2	Forceps
Severe metabolic acidosis	3	14	3	0	12	5	0	C/S
	6	17	0	0	16	0	1	C/S
	30	0	13	4	0	15	2	No Action
	41	3	9	5	1	13	3	No Action
Acidosis	7	7	0	10	2	3	12	No Action
	26	0	10	7	0	11	6	No Action
	28	1	14	2	0	16	1	Forceps
	35	17	0	0	17	0	0	C/S
	38	15	0	2	16	0	1	No Action

Table 5.23: Summary of the experts' and system's actions in cases with poor outcome.

The experts' and the system's management was considered for each group separately and is presented in tables 5.24, 5.25 and 5.26.

Experts	Number of cases where intervention was recommended (max 3)	
	Review 1	Review 2
A	2	1
B	3	3
C	2	2
D	2	2
E	2	2
F	2	2
G	2	2
H	3	3
I	2	2
J	2	2
K	2	1
L	3	2
M	2	2
N	2	2
O	2	2
P	2	2
Q	2	2
System	2	2
Clinical	1	

Table 5.24: Operative intervention in the cases with birth asphyxia (3 cases).

Experts	Number of cases where intervention was recommended (max 4)	
	Review 1	Review 2
A	3	4
B	3	4
C	4	4
D	4	4
E	4	4
F	2	4
G	4	4
H	4	4
I	3	4
J	3	4
K	4	3
L	4	4
M	3	3
N	4	3
O	4	4
P	3	2
Q	3	3
System	2	2
Clinical	2	

Table 5.25: Operative intervention in cases with significant metabolic acidosis (4 cases).

Experts	Number of cases where intervention was recommended (max 5)	
	Review 1	Review 2
A	1	2
B	3	3
C	4	5
D	5	4
E	5	4
F	4	4
G	5	4
H	5	4
I	3	4
J	5	3
K	3	2
L	5	5
M	2	4
N	4	4
O	5	5
P	3	4
Q	2	4
System	2	2
Clinical	2	

Table 5.26: Operative intervention in the cases with acidosis (5 cases).

Summary of results in cases with poor perinatal outcome.

The system.

1. The system recommended necessary operative intervention in the same number of cases with birth asphyxia as the majority of experts (2 out of 3).
2. The system consistently recommended operative delivery in one case of birth asphyxia which received no intervention in clinical practice.
3. The system was within the range of the experts' operative interventions in the significant metabolic acidosis and acidosis groups.
4. The system recommended no action in 2 cases (cases 41 and 30), which resulted in significant metabolic acidosis, whereas the majority of experts consistently recommended intervention during the second stage of labour.
5. The system did not recommend operative intervention in 2 cases which resulted in acidosis (cases 26 and 38). The majority of experts recommended operative intervention in case 38 by CS and intervention during the second stage in case 26.
6. Both the system and the majority of experts did not recommend intervention in case 15 which resulted in birth asphyxia.

The cases where the system recommended no action are discussed in 5.5.8.

The experts.

1. 2 experts (B and H) consistently recommended operative intervention in all cases with birth asphyxia.
2. In one of their reviews, 2 experts (A and K) failed to detect 2 of the 3 cases of birth asphyxia.
3. Expert H obtained a remarkable performance by recommending operative intervention in all 12 cases which resulted in poor outcome in her first review, and all but one case (with simple acidosis) in her second.

5.5.7 Intervention in cases with good clinical outcome.

There were 11 cases which obtained good outcome (cord artery pH > 7.20) after a normal vaginal delivery. Operative intervention in these cases was unnecessary. These cases were investigated to establish whether either the system or the experts recommended unnecessary operative intervention. Table 5.27 summarises these cases and indicates for each case, the experts' and the system's recommended actions.

Outcome	Case Number	Review 1			Review 2			System Recommendation
		No. of experts recommending			No. of experts recommending			
		C/S	Forceps	No Action	C/S	Forceps	No Action	
Good	12	3	2	12	2	9	6	No action
	14	1	6	10	0	5	12	No action
	21	2	1	14	1	1	15	No action
	22	0	0	17	0	0	17	No action
	31	0	0	17	0	0	17	No action
	33	0	2	15	0	0	17	No action
	39	0	0	17	0	1	16	No action
	46	0	3	14	0	1	16	No action
	47	0	0	17	0	0	17	No action
	49	0	0	17	0	0	17	No action
	50	0	1	16	0	1	16	No action

Table 5.27: Intervention in cases with good clinical outcome.

The level of each reviewers operative intervention in these cases is shown in table 5.28.

Expert	No. of cases where operative intervention was recommended			
	Review 1		Review 2	
	C/S	Forceps	C/S	Forceps
A	0	0	0	1
B	2	1	1	2
C	0	1	0	3
D	0	1	0	1
E	0	1	1	0
F	1	0	0	1
G	0	1	0	1
H	1	0	1	2
I	0	0	0	1
J	0	3	0	1
K	0	1	0	1
L	0	2	0	1
M	0	0	0	0
N	0	0	0	1
O	0	1	0	1
P	0	0	0	0
Q	2	0	0	1
System	0	0	0	0

Table 5.28: Each reviewers unnecessary operative intervention.

Summary of results in cases with good perinatal outcome.

The system.

1. The system recommended no unnecessary intervention in cases with good outcome.

The experts.

1. Experts M and P consistently recommended no operative intervention in cases with good outcome.
2. Experts, H and Q each recommended 2 unnecessary CSs. Expert H also recommended 2 unnecessary second stage interventions and expert Q recommended 1 unnecessary second stage intervention.
3. Expert B recommended 3 unnecessary CS deliveries and 3 unnecessary interventions during the second stage.

5.5.8 Case discussion.

Four tests have identified that the management recommended by the system was different (not necessarily worse) to the majority of experts in 14 cases. In addition, the system, along with the majority of experts, did not identify a case which resulted in birth asphyxia (case 15). These cases are discussed in detail here and are summarised in table 5.29 along with the identifying tests. **A general discussion of the system's and experts' performance is presented in chapter 7.**

1. Analysis of the levels of agreement obtained compared to the experts.
2. Cases recommended for caesarean section (CS).
3. Cases with poor clinical perinatal outcome, split into three groups (birth asphyxia, metabolic acidosis and acidosis).
4. Fetal blood sampling rates (FBS).

Test	Identified Cases														
	2	8	11	15	16	17	24	26	30	32	37	38	41	44	48
ANOVA		•	•									•		•	•
CS	•				•		•			•	•	•		•	
Birth asphyxia				•											
Metabolic acidosis									•				•		
Acidosis								•				•			
FBS						•									

Table 5.29: Cases identified for discussion.

The 15 identified cases were investigated to determine whether the systems management was in any way inappropriate.

Cases with poor outcome.

Case 15 - Birth asphyxia.

The system consistently did not recommend intervention in this case which resulted in birth asphyxia. However, nor did 14 of the experts. The mother involved made excellent progress through labour; her cervix dilated from 3cm to 10 cm (fully) in approximately 4 hours. No fetal blood samples were obtained during the labour but from an interpretation of the CTG and knowing the outcome, two experienced clinicians considered that this baby started to become compromised during the second stage of labour. When second stage began, the arterial pH was adjudged to be approximately 7.20. After a second stage lasting 45 minutes, the baby was delivered with birth asphyxia. This is an extremely unusual case which has been reviewed during a meeting of international physiologists and various study groups. The heart rate changes were very subtle. The fact that only two experts consistently recommended intervention shows that it would have been difficult to prevent a poor outcome in this case. This baby did however recover satisfactorily.

Cases 26 30, 38 and 41.

None of these cases had operative intervention clinically. In cases 30 and 38 there occurred a gross baseline change from a slight tachycardia (> 160 bpm) to a bradycardia between 95 and 100 bpm which prompted the experts to recommend immediate intervention. The system identified these heart rate changes but was not seriously concerned (score 2) as the rule for action under these circumstance requires the baseline to fall below 90 bpm. In cases 26 and 41, a severe bradycardia occurred (baseline < 90) in the second stage of labour for which the system considered recommending immediate emergency delivery (score 4). This was the point where the experts recommended their interventions. Delivery occurred in these cases, very soon after the heart rate changes.

It can be noted from these cases which obtained a poor outcome clinically that the system;

1. Did not react as quickly during the second stage of labour as the experts; the system does not make a distinction between the first and second stage in terms of the time it allows abnormalities to persist before recommending intervention.
2. The system did not interpret a rapid baseline fall from slight tachycardia to slight bradycardia as a severely abnormal event. The experts did and identifies a gap in the system's knowledge.

Cases where the majority of reviews recommended delivery by CS but the system recommended either no action or a later delivery by forceps.

Case 2.

There was significant expert disagreement in the management of this case. This was due to the considerable variation in the recommended timing of CS. The average timing was in the 23rd segment, the first was in segment 2 (expert H in both reviews) and the final recommendation was in segment 44. Five FBS were obtained clinically throughout the labour but none were less than 7.24 after which this baby was delivered by CS. The reasons given clinically were a falling fetal scalp pH combined with poor progress in labour. The outcome at birth was entirely normal with cord artery pH of 7.32, BDecf of 1 and 1 minute Apgar score 9. The system also requested 5 FBS but because the pH never fell below the normal range it did not recommend operative intervention.

The lady in this case made poor progress through labour which, coupled with an abnormal CTG, was enough to lead the experts to deliver by CS. The system's recommendations seem reasonable because there was no evidence of fetal compromise.

Case 16

There was a high variation in the recommended timing of CS by the experts. The averaged timing was in the 35th segment with the first recommended in the 16th segment and the last in the 47th. The system identified the CTG as abnormal and recommended 5 FBS. However, as each result was normal the system acknowledged that the CTG was abnormal but as the pH was normal and stable it decided to recommend no further FBS nor intervention. The period between the last recommended FBS and the end of recording was 4 hours 15 minutes which could be considered as too long under these circumstances. Clinically, a normal FBS result of 7.25 was obtained 1 hour before delivery by CS. The decision to deliver by CS was taken because of the fetal scalp pH and poor progress in labour. The cord artery pH was 7.05 and BDecf was 7 which is close to the definition of acidosis (pH < 7.05). The Apgars were 7 and 9 at 1 and 5 minutes.

Case 24

There was good expert agreement for delivery by CS in this case. 16 experts in 30 reviews (max 17 experts and 34 reviews) recommended a CS in the 9th segment of a CTG 10 segments long. The system was considering operative delivery (score 4) by CS on the 9th and 10th segment. Clinically, the decision was taken to deliver by CS because the CTG was abnormal early in labour with the cervix just 4cm dilated. The cord artery pH was 7.17 with a BDecf of 2 and Apgars 6 and 9. The system's interpretation was similar to the experts and if this CTG had continued in a similar fashion, the system would have recommended CS.

Case 32

There was good agreement amongst the experts in this case with 14 experts in 26 reviews recommending delivery by CS on average in the 14th segment. However, 3 experts did not recommend CS in either of their reviews and a further 2 experts recommended a CS in one review. The system had identified a deterioration in the trace reaching a protocol score of 2 for the final two segments (16 and 17) but did not recommend intervention. Clinically the baby was delivered by CS with cord artery pH of 7.15, BDecf of 10 and Apgars 6 and 9 at 1 and 5 minutes. The decision was taken because of a low scalp pH early in labour (cervix 4cm).

Case 37

Most experts recommended CS and did so in two distinct timings which coincide with two episodes of bradycardia from which the fetus made a good recovery both times. The first period of bradycardia occurred in segments 35 and 36 which caused 6 experts in 7 reviews to recommend CS. The second episode occurred in segments 50 and 51 and led to the remaining experts who recommended a CS to do so. The system recognised both events as very abnormal and reached a protocol score of 4 on both occasions which indicated that it was considering delivery. But as the trace recovered to satisfactory, no further action was recommended. One and a half hours after this second event the trace again deteriorated albeit to a less serious extent than previously. Because of the previous episodes, the system recommended immediate delivery, which because the cervix was fully dilated was by forceps. This was also the view in 5 expert reviews. Clinically too, a forceps delivery took place. The cord arterial pH was 7.19 and BDecf was 7 and Apgars were 8 and 9 at 1 and 5 minutes. With such an outcome, it could reasonably be argued that the system managed this case more effectively than the majority of experts who recommended CS.

Case 44

All experts recommended intervention in this case close to the end of the recording when in clinical practice a CS took place. The system had been registering slight concerns but these were not sufficient to recommend intervention. The system's recommendations were different from the experts but it is difficult to say which were more appropriate as the clinical outcome was not poor (cord artery pH of 7.11)

ANOVA analysis by ranks.

Apart from cases 38 and 44 previously discussed, the ANOVA test identified 3 further cases where the system had significantly lower agreement than the experts.

Case 8.

The median agreement the system scored with the experts was 65%. There were only minor differences between the system's management and the experts'. Most experts requested an

FBS around segments 7, 8 and 9 whereas the system recommended no intervention. In the final segment, where clinically a forceps delivery was performed, the system was considering immediate delivery (score 4). Most of the experts also did not recommend intervention. The clinical outcome was good and so the system's management was appropriate.

Case 11.

The median agreement with the experts was 65%. The system recommended a FBS in segment 2 but no expert did. However, several did so later. The system considered immediate delivery (score 4) in the final segment and several of the experts recommended intervention. The baby was actually delivered normally with a satisfactory outcome and the system's management was appropriate.

Case 48.

The median agreement with the experts was 50%. A severe bradycardia occurred in this case in segments 6 and 7. The system responded by considering immediate delivery (score 4) as did most experts. Two expert reviews recommended immediate CS. In segments 25 - 28 the CTG again became abnormal. The system recommended a CS in segment 29 because of the deteriorating CTG and the fact that a severe bradycardia had previously occurred. Most experts recommended delivery by CS some 1 hour after the system.

FBS test

Case 17.

The system was inconsistent in its management of this case. In segment 13, most experts recommended a CS. At this time, the system recognised the abnormal CTG and asked whether this baby was considered high risk. The operator in the first review replied No to which the system recommended a FBS followed immediately by CS. The operator in the second review replied Yes and the system recommended immediate CS without requesting a FBS. This inconsistency came from the fact that some time before, the CTG showed abnormal heart rate changes. The obstetrician considered these significant and so considered the baby to be at high risk whereas the engineer did not. The difference between the system's decisions to deliver by CS was 2 segments and both decision points were within the range of the experts.

Chapter 5 summary.

Chapter 5 described the evaluation and validation of an intelligent system to assist in the management of labour.

The formulated expert knowledge was implemented in an expert system and then evaluated in a study which compared the management of the system with the collaborating expert in a retrospective review of 31 labours. This study found that the system compared favourably with the expert; compared to clinical practice, both would have reduced fetal blood sampling (50%) and unnecessary operative intervention. In addition, both recommended earlier intervention in the same 4 cases which may have improved outcome. This study demonstrated that expert knowledge could be formulated for this difficult task. This work was presented at an international conference and published in full in the refereed conference proceedings (Ifeachor et al, 1990).

The automatic feature extraction methods were combined with the expert system. An evaluation of the full system was undertaken to compare the recommendations of the system with 3 experienced obstetricians from Plymouth. This study found that the system obtained good agreement with the clinicians; every action recommended by the system was also recommended by at least one other clinician. This study concluded that not only could the system agree with the collaborating expert, but in these cases, it could also agree with the management of 2 other clinicians not previously involved in the development of the system. This work has formed the basis of a journal paper which has been accepted for publication (Keith et al, 1993).

The possible limitation of a single centre developing this type of system in isolation was recognised. 17 experienced clinicians from 16 leading centres in fetal monitoring within the UK, took part in the validation of the system. The CTGs and associated clinical data of 50 cases with a range of outcomes were selected from a database of 2400 high risk labours. These 50 cases contained those considered to be most difficult to interpret. Each case was reviewed by the experts twice, one month apart to establish their consistency. Each CTG was contained in an opaque box and revealed in 15 minute segments. The relevant obstetric history was made known before each case was reviewed and any relevant events during the labour were made known at the appropriate time. Each reviewer scored every 15 minute segment according to a protocol to reflect the concern they had for the fetus. The task for each expert was to manage each labour with minimal intervention without jeopardising the safety of the fetus. The system reviewed the cases twice with 2 different operators working independently; one obstetrician and one engineer. A method was devised to measure the agreement in terms of labour management between any two sequences of scores recorded for a case. This derivation is presented in chapter 6. Kappa statistics which measure

agreement in excess of that expected by chance, together with the Kruskal-Wallis one way analysis of variance by ranks were used to analyse the agreements.

It was found that;

The experts.

1. All experts obtained an agreement with each other significantly better than could have been expected by chance. The average inter-agreement was 70%.
2. All experts were consistent significantly in excess of that which could have been expected by chance. The average intra-agreement was 80%.
3. There was significant ($\alpha = 0.1$) inter-expert disagreement in just 1 case (case 2).
4. In the cases recommended for CS by 14 experts, at least 24 of the other 32 expert reviews also recommended delivery by CS.
5. In the 31 cases where at least one expert recommended a CS, the majority of other experts who also recommended delivery by CS did so within ± 1 segment (15 minutes) and two thirds did so within ± 2 segments.
6. The majority of experts recommended operative intervention in 10 out of the 12 cases with acidosis ($\text{pH} < 7.05$) at delivery.
7. The majority of experts did not recommend operative intervention in cases which obtained a normal delivery with good outcome, except in 1 case in the second review where 9 out of 17 experts recommended intervention in the second stage.

The system.

1. The system agreed with the experts well and significantly in excess of chance (67%, $\alpha = 0.01$, $p < 0.0000003$).
2. In 45 cases, the level of agreement the system achieved with the experts was not significantly lower than the experts achieved with each other.
3. The actions recommended by the system were highly consistent (99%, $\alpha = 0.01$), even when used by someone with no labour ward experience.
4. The system recommended delivery by CS in 11 cases. On average more than 31 of the 34 expert reviews also recommended CS delivery in these cases. The majority of these experts did so within ± 1 segment (15 minutes) of the system's recommendation and two thirds did so within ± 2 segments (30 minutes).
5. The system identified as many of the birth asphyxiated cases as the majority of experts and 1 more than was acted upon clinically.
6. The system recommended no unnecessary intervention in the cases with good outcome (spontaneous vaginal delivery and $\text{pH} > 7.20$) which was better than all but two of the experts.

The findings from the validation study were both positive and conclusive. The experts were able to agree and be consistent in the management of labour and the system was able to obtain a performance which made it indistinguishable from the experts.

Chapter 5 references.

- Ifeachor E C, Keith R D F, Westgate J and Greene K R. (1991) An expert system to assist in the management of labour. *Expert Systems World Congress Proceedings*. Orlando Florida December 16th-19th Liebowitz J. (ed). Pergamon Press, Oxford UK 4: 2615-2622.
- Keith R D F, Westgate J, Hughes G W, Ifeachor E C and Greene K R. (1993) Preliminary evaluation of an intelligent system for the management of labour. *J Perinat Med*. (Accepted for publication)
- O'Keefe R M, Balci O and Smith E P. (1987) Validating expert system performance. *IEEE Expert*. Winter, 81-90.
- O'Leary T J, Goul M, Moffitt K E and Radwan AE. (1990) Validating expert systems. *IEEE Expert*. June, 51-58.

Chapter 6

Derivation of a measure of agreement and the application of statistical hypothesis tests.

- 6.1 Introduction 181
- 6.2 Derivation of a measure to assess agreement in labour management 181
- 6.3 Derivation of the statistical distribution of agreement 192
- 6.4 Statistical tests to investigate agreement 199
- 6.5 Application of statistical hypothesis tests 205

Chapter 6 summary215

Chapter 6 references 216

6.1 Introduction.

This chapter presents the mathematical derivation of the measure used to assess the agreement between the sets of scores recorded during the validation study (chapter 5, section 5.4). In addition, the application of statistical tests which consider various hypotheses regarding the significance of the measured agreements is presented. The results involving this measure have been previously summarised in chapter 5 (section 5.5.2) but are reported in detail here.

6.2 Derivation of a measure to assess agreement in labour management.

The scores that each reviewer recorded for a given case represents how they would have managed the labour. The scores obtained for case 1 are shown in chapter 5, table 5.6 and the complete set of scores for the 50 cases are given in appendix H.

A single measure was required to indicate the agreement between any two given reviewers recommendations. Put formally, a measure of the agreement between 2 time dependent, discrete sequences, $A(t)$ and $B(t)$, was required such that when they were identical they scored a maximum agreement of 1.0 and when there was no similarity, they scored 0. If the length of each sequence was N then $A(t)$ and $B(t)$ can be represented by equation 6.1.

$$\begin{aligned} A(t) &= [a(1), a(2), a(3), \dots, a(N)] \\ B(t) &= [b(1), b(2), b(3), \dots, b(N)] \end{aligned} \tag{6.1}$$

A measure of the agreement between $A(t)$ and $B(t)$, termed the cross agreement coefficient, $Cac(A,B)$, would simply be given by equation 6.2.

$$Cac(A, B) = \sum_{t=1}^N g(t) \quad \text{where } g(t) = \begin{cases} 1, & \text{when } a(t) = b(t) \\ 0, & \text{otherwise} \end{cases} \tag{6.2}$$

If the values $a(t)$ and $b(t)$ were in the range, 1 to s say, then this same measure of agreement could be obtained by representing the relationship between $a(t)$ and $b(t)$ in an agreement matrix, $M(A,B)$, such that,

		b(t)						
		1	2	3	.	s		
M(A,B) =	a(t)	1	0	0	0	0	0	
		2	0	1	0	0	0	0
		3	0	0	1	0	0	0
		.	0	0	0	1	0	0
		.	0	0	0	0	1	0
		s	0	0	0	0	0	1

Here, the cross agreement coefficient would be given by the sum of the individual mappings of a(t) and b(t) represented by equation 6.3.

$$Cac(A, B) = \sum_{t=1}^N M_{a(t), b(t)} \quad t = 1, 2, 3, \dots N. \quad (6.3)$$

However, this expression is dependent not only on the similarities between A(t) and B(t), but also on the sequence length, N. A measure independent of the sequence length could be obtained if the cross agreement coefficient were normalised to the maximum agreement score possible. This is given when A(t) = B(t) and would equal the total number of scores, N. This maximum agreement could also be obtained using the agreement matrix with the substitution A(t) = B(t).

		a(t)						
		1	2	3	.	s		
M(A,A) =	a(t)	1	0	0	0	0	0	
		2	0	1	0	0	0	0
		3	0	0	1	0	0	0
		.	0	0	0	1	0	0
		.	0	0	0	0	1	0
		s	0	0	0	0	0	1

This maximum agreement measure will be termed the auto agreement coefficient, Aac(), and for A(t), this is given by equation 6.4.

$$Aac(A) = Cac(A, A) = \sum_{t=1}^N M_{a(t), a(t)} \quad (6.4)$$

For the method developed so far, $Cac(A, B) = Cac(B, A)$ and $Aac(A) = Aac(B)$, and the normalised agreement between A(t) and B(t) is given by equation 6.5.

$$Agreement(A, B) = \frac{Cac(A, B)}{Aac(A)} = \frac{\sum_{t=1}^N M_{a(t), b(t)}}{\sum_{t=1}^N M_{a(t), a(t)}} \quad (6.5)$$

Now consider the situation where the lengths of A(t) and B(t) are allowed to differ. Here, $Aac(A) \neq Aac(B)$. A combined measure of the auto agreement would be given by, $Aac(A) + Aac(B)$ but for the situation where the sequences were identical, the combined auto agreement would be a factor of 2 greater than the cross-agreement and would consequently limit the measure to a maximum of 0.5. This could be overcome by equation 6.6.

$$Agreement(A, B) = \frac{2 \cdot Cac(A, B)}{Aac(A) + Aac(B)} = \frac{2 \sum_{t=1}^N M_{a(t), b(t)}}{\sum_{t=1}^N M_{a(t), a(t)} + \sum_{t=1}^P M_{b(t), b(t)}} \quad (6.6)$$

Where, N is the length of sequence A(t) and P is the length of sequence B(t).

Equation 6.6 is now valid whilst $Cac(A, B) = Cac(B, A)$. A more general form of the equation which does not depend on this could be obtained if the combined cross agreement coefficient were represented by $Cac(A, B) + Cac(B, A)$. Substituting this gives equation 6.7.

$$Agreement(A, B) = \frac{Cac(A, B) + Cac(B, A)}{Aac(A) + Aac(B)} = \frac{\sum_{t=1}^N M_{a(t), b(t)} + \sum_{t=1}^P M_{b(t), a(t)}}{\sum_{t=1}^N M_{a(t), a(t)} + \sum_{t=1}^P M_{b(t), b(t)}} \quad (6.7)$$

6.2.1 Derivation of a weighted agreement matrix.

In the development of the agreement measure so far, a relationship between a(t) and b(t) was assumed which gave rise to the identity agreement matrix. But this relationship may not be appropriate for all applications. Consider, for example the scores obtained in the validation study. The protocol was restricted to allow the experts to record a score in the range 1 to 5 for each 15 minute segment which gives the following identity matrix,

		b(t)				
		1	2	3	4	5
M(A,B) =	a(t)	1	0	0	0	0
		2	1	0	0	0
		3	0	1	0	0
		4	0	0	1	0
		5	0	0	0	1

It was found that each expert scored an average of 2129 segments. But as expected, these scores were not evenly distributed; in fact, on average the experts attached a score of 1 or 2

(which translates to not intervening in the labour) to 87.3% of the segments reviewed. If the identity matrix were used, the agreement measure would be biased in favour of the experts because they scored the large majority of segments just two of five values. In addition, the identity matrix assumes that an agreement between two experts scoring 5 (emergency CS) is as significant as when they both score a 1 (non-intervention), when clearly the implications are profoundly different. Therefore, this matrix does not adequately represent the specific case of labour management.

An appropriate matrix could be derived according to some consensus obtained with clinicians but such an approach would justifiably be open for criticism because it is entirely subjective. An alternative, more objective approach must be found.

Digital signal processing approach.

Correlation is a very important technique used in signal processing to measure the similarities between two signals. Here, the correlation between two sequences with zero lag, is given by the sum of the products of the corresponding pairs of points.

$$r_{AB} = \sum_{t=1}^N a_t \cdot b_t$$

This process can be represented in the agreement matrix as,

		b(t)				
		1	2	3	4	5
M(A,B) =	1	1	2	3	4	5
	2	2	4	6	8	10
	3	3	6	9	12	15
	4	4	8	16	20	24
	5	5	10	15	20	25

However, this matrix is also unsuitable for our specific application. For example, the agreement given when $a(t) = 1$ and $b(t) = 5$ (and vice versa) is 5. This combination of scores represents the widest clinical difference, but with this matrix, they would represent a greater agreement than when the scores were identical and both equal to 1 or 2.

Cost benefit analysis approach.

A different approach would be to consider the cost implications of certain actions. The financial costs involved in managing labour have been calculated (Clark, 1991); to record the CTG costs £21.00 and to obtain a FBS costs a further £22.00. To intervene with

forceps could cost up to £773.00 and by CS, £1560.00. However the implications for the patient (mother and fetus) are also important but would be more difficult to quantify.

A probabilistic approach.

The problem with the identity matrix is that it rewards equally an agreement for scores of 1 which is bound to occur often, as it does for scores of 5 which occur rarely. To expand, the protocol instructed that a case review be terminated when a score permitting delivery was reached (5 and under certain circumstances 3). Consequently, the maximum number of 5's an expert could have recorded in their 2 reviews of 50 cases was 100. On the other hand, an expert could have recorded a score of 1 for every possible segment which would amount to a maximum total of 2376. Clearly, a score of 5 is far more significant than a score of '1' and implies the agreement matrix should be biased towards the less common scores. Table 6.1 shows the combined frequency distribution of the experts' recorded scores.

Score	1	2	3	4	5	Total
Frequency	15749	15864	2562	1427	607	36,209
Probability	0.43496	0.43812	0.07076	0.03941	0.01676	1
Significance	2	2	14	25	60	

Table 6.1: Combined frequency distribution of the experts scores.

These frequencies show the bias of the scoring towards the lower scores. This bias could be equalised if the significance of each score were found. The probabilities of the different scores, $a(t)$, are given by equation 6.8 and shown in table 6.1.

$$p\{a(t)\} = \frac{\text{frequency of score}(a)}{\text{Total number of scores}} = \frac{\sum a}{N} \quad (6.8)$$

The significance of each score was taken as the approximate inverse of the probability and results in the following agreement matrix,

		b(t)				
		1	2	3	4	5
M(A,B) =	1	2	0	0	0	0
	2	0	2	0	0	0
	3	0	0	14	0	0
	4	0	0	0	25	0
	5	0	0	0	0	60

These figures were discussed with clinicians and they agreed that this agreement matrix represented a fair reflection of the clinical implications of each protocol score.

6.2.2 Measuring partial agreement.

The partial agreement between different scores.

The matrix derived so far only attributes an agreement to those segments scored identically. However, the difference between two different scores given for the same segment may not represent complete agreement, but it may not represent complete disagreement either. This is currently the case and indicated by the zero values given to the off-diagonal mappings. For certain combinations of scores it may be appropriate to attribute a value greater than zero to indicate a partial agreement. These values would of course be subjective, but they should be reasonable to reflect the true partial agreement that exists. Two conditions were imposed before the partial agreements were derived.

1. Given the scores, $a(t)$ and $b(t)$, recorded for segment, t ; if $a(t) \neq b(t)$ then the partial agreement should be less than when $a(t) = b(t)$.
2. The agreement matrix should be symmetrical about the agreement diagonal, $M\{a(t) = b(t)\}$ such that the partial agreement, $M_{a(t)b(t)} = M_{b(t)a(t)}$.

With these conditions in place, consider the partial agreements, where $a(t) \neq b(t)$ for reviewers A and B.

Partial agreements involving a score of 1.

- The agreement given to $a(t) = 1$ and $b(t) = 2$ should be high. If reviewer A scored a labour throughout, 1, and B scored it throughout, 2, then clinically, the labours were managed similarly as no interventions were recommended. Condition 1 requires a partial agreement should be < 2 , so let it be 1.5.
- No other partial agreement involving a score of 1 seems appropriate as these (3, 4, and 5) involve intervention or a high level of concern.

These assignments are represented in the following agreement matrix,

$$M(A,B) = \begin{array}{c|ccccc} & & \text{b(t)} & & & & \\ & & 1 & 2 & 3 & 4 & 5 \\ \text{a(t)} & 1 & 2 & 1.5 & 0 & 0 & 0 \\ & 2 & 1.5 & 2 & & & \\ & 3 & 0 & & 14 & & \\ & 4 & 0 & & & 25 & \\ & 5 & 0 & & & & 60 \end{array}$$

Partial agreements involving a score of 2.

- A partial agreement with $a(t) = 2$ and $b(t) = 3$ is appropriate as both recognise some concern. But this should be less than when $a(t) = 1$ and $b(t) = 2$, as the clinical difference is greater. If the partial agreement should be < 1.5 , let it be 1.
- No other partial agreement involving a score of 2 seems valid as these (4 and 5) involve very high levels of concern.

$$M(A,B) =$$

		b(t)				
		1	2	3	4	5
a(t)	1	2	1.5	0	0	0
	2	1.5	2	1	0	0
	3	0	1	14		
	4	0	0		25	
	5	0	0			60

Partial agreement involving a score of 3.

- A measure of agreement for $a(t) = 3$ and $b(t) = 4$ is reasonable as both indicate significant levels of concern. This must be less than 14, so let it be 10.
- There is also some agreement in $a(t) = 3$ and $b(t) = 5$ as both represent intervention. However, these scores represent the difference between an intervention to obtain further information and immediate emergency delivery. As such, the partial agreement should be relatively low and less than the partial agreement given for $a(t) = 3$ and $b(t) = 4$. Let this partial agreement be 5.

$$M(A,B) =$$

		b(t)				
		1	2	3	4	5
a(t)	1	2	1.5	0	0	0
	2	1.5	2	1	0	0
	3	0	1	14	10	5
	4	0	0	10	25	
	5	0	0	5		60

Partial agreement involving a score of 4.

- A measure of agreement for, $a(t) = 4$ and $b(t) = 5$ is reasonable as both scores relate to emergency delivery. This should be less than 25, so let this agreement be 20.

$$M(A,B) =$$

		b(t)				
		1	2	3	4	5
a(t)	1	2	1.5	0	0	0
	2	1.5	2	1	0	0
	3	0	1	14	10	5
	4	0	0	10	25	20
	5	0	0	5	20	60

Partial agreement in the timing of decisions.

The agreement matrix permits a partial agreement between two different scores given for the same segment. In certain situations, for a given segment score $a(t)$, there may also exist a partial agreement in the adjacent scores $b(t-1)$, $b(t+1)$, $b(t-2)$, $b(t+2)$, etc. To illustrate the point consider example 6.1 which examines hypothetical scores which could have been obtained in the study.

Example 6.1.

	Segment					
	1	2	3	4	5	6
A(t)	2	2	2	3	5	
B(t)	1	2	2	2	3	5

Using equation 6.7,

$$Agreement(A,B) = \frac{Cac(A,B) + Cac(B,A)}{Aac(A) + Aac(B)} = \frac{\sum_{t=1}^N M_{a(t),b(t)} + \sum_{t=1}^P M_{b(t),a(t)}}{\sum_{t=1}^N M_{a(t),a(t)} + \sum_{t=1}^P M_{b(t),b(t)}}$$

with the statistically weighted matrix gives,

$$Aac(A) = 2+2+2+14+60 = 80$$

$$Aac(B) = 2+2+2+2+14+60 = 82$$

$$Cac(A,B) = 1.5+2+2+1+5+0 = 11.5$$

$$Cac(B,A) = 1.5+2+2+1+5+0 = 11.5$$

and substituting gives,

$$agreement(A,B) = \frac{11.5+11.5}{80+82} = 0.142 \text{ or } 14.2\%$$

This low agreement measure does not reasonably reflect the closeness of the clinical actions. Expert A and expert B have both been concerned for the fetus which promoted them to recommend a FBS immediately followed by a CS. The experts did not recommend these actions at the same time, but they did recommend them just 1 segment (15 minutes) apart. If either expert's recommendations had been adopted clinically, the patient would have been managed very similarly indeed which is not reflected with an agreement of 14.2%.

Example 6.1 makes the case that a partial agreement should be allowed between $a(t)$ and $b(t-1)$, $b(t+1)$ etc. (and vice versa). A measure which allowed partial temporal agreement between 2 sequences, $A(t)$ and $B(t)$ at time t would be obtained by equation 6.9.

$$\max \{w(0)M_{a(t)b(t)}, w(1)M_{a(t)b(t+1)}, w(1)M_{a(t)b(t-1)}, \dots, w(+T)M_{a(t)b(t+T)}, w(-T)M_{a(t)b(t-T)}\} \quad (6.9)$$

which can be written concisely as equation 6.10.

$$\max_{\tau=0}^{\tau=T} \{W(t-\tau) \cdot M_{a(t)b(t-\tau)}\} \quad \text{for } \tau = 0, \pm 1, \pm 2, \pm 3, \dots, \pm T \quad (6.10)$$

Equation 6.10, effectively allows individual scores to shift by τ sample points in either direction to find the position of maximum agreement. A weighting function, $W(t-\tau)$, was incorporated to reduce the partial temporal agreement as the relative time between similar scores increases. It was decided that no partial agreement should be given for scores ≥ 1 hour apart (± 4 segments). The choice of weighting function is clearly arbitrary but a cosine function was considered most appropriate. This function is given by equation 6.11.

$$W(t-\tau) = \cos \frac{\pi \tau}{8} \quad \text{for } \tau = 0, \pm 1, \pm 2, \pm 3 \quad (6.11)$$

and is illustrated in figure 6.4.

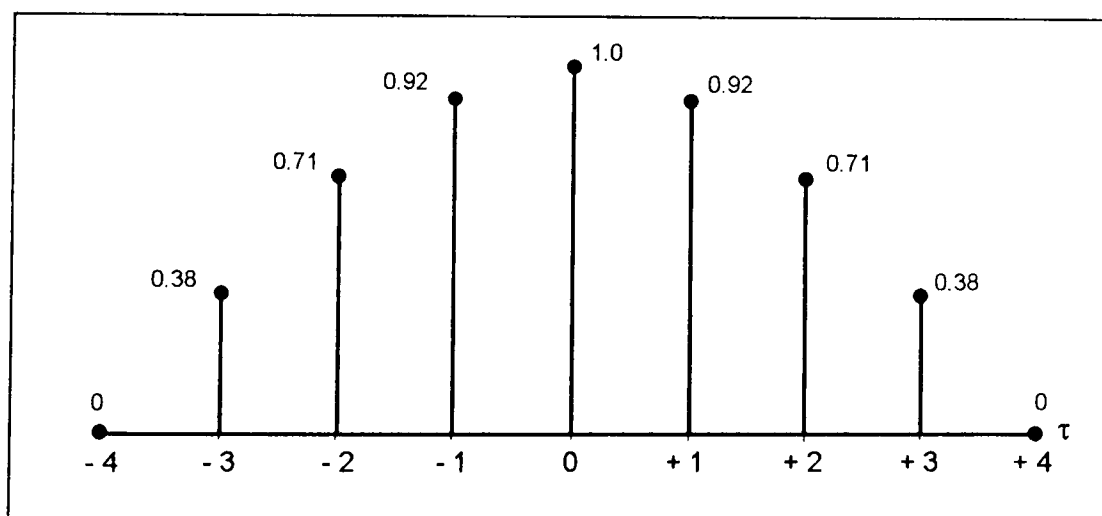


Figure 6.4: The window function used for temporal partial agreement.

The suitability of this window was discussed with clinicians and considered reasonable as it gives a high temporal partial agreement for $\tau = \pm 1$, and rolls off at a desirable rate to 0.

Substituting this into equation 6.3 obtains the cross agreement function between A(t) and B(t) as equation 6.12.

$$Cac(A, B) = \max_{\tau=-3}^{\tau=+3} \left\{ W(t - \tau) \cdot \sum_{t=1}^N M_{a(t), b(t-\tau)} \right\} \quad (6.12)$$

Substituting this into equation 6.7 obtains the agreement equation 6.13.

$$Agreement(A, B) = \frac{\max_{\tau=-3}^{\tau=+3} \left\{ w(t - \tau) \cdot \sum_{t=1}^N M_{a(t), b(t-\tau)} \right\} + \max_{\tau=-3}^{\tau=+3} \left\{ w(t - \tau) \cdot \sum_{t=1}^P M_{b(t), a(t-\tau)} \right\}}{\sum_{t=1}^N M_{a(t), a(t)} + \sum_{t=1}^P M_{b(t), b(t)}} \quad (6.13)$$

This equation retains the desirable properties of equation 6.7; It has a maximum value of 1.0 for identical sequences which will always be the case so long as the maximum weighting in the weighting function, $W(t - \tau)$, is 1.0.

Example 6.2.

Consider the operation of equation 6.13 for the same sequences considered in example 6.1.

	Segment					
	1	2	3	4	5	6
A(t)	2	2	2	3	5	
B(t)	1	2	2	2	3	5

The calculation of the auto agreement coefficients remain unchanged,

$$Aac(A) = 2 + 2 + 2 + 14 + 60 = 80$$

$$Aac(B) = 2 + 2 + 2 + 2 + 14 + 60 = 82$$

The cross agreement coefficient $Cac(B, A)$ is given by,

$$\begin{aligned} Cac(A, B) = & \max\{1.0*1.5, 0.92*2, 0.71*2, 0.38*2\} + \\ & \max\{1.0*2, 0.92*1.5, 0.92*2, \dots\} + \\ & \max\{1.0*2, 0.92*2, 0.92*1, 0.71*1.5, 0.71*1, \dots\} + \\ & \max\{1.0*1, 0.92*1, 0.92*14, 0.71*2, 0.71*5\} + \\ & \max\{1.0*5, 0.92*0, 0.92*60, \dots\} \end{aligned}$$

$$Cac(A, B) = 0.92*2 + 1.0*2 + 1.0*2 + 0.92*14 + 0.92*60$$

$$\underline{Cac(A, B) = 73.92}$$

Similarly,

$$Cac(B, A) = 0.92*1.5 + 1.0*2 + 1.0*2 + 0.92*2 + 0.92*14 + 0.92*60$$

$$\underline{Cac(B, A) = 75.42}$$

and,

$$Agreement \{A(t), B(t)\} = \frac{Cac(A, B) + Cac(B, A)}{Aac(A) + Aac(B)} = \frac{73.92 + 75.42}{80 + 82} = 0.922 = 92.2\%$$

This measure of 92.2% reflects a more appropriate measure of agreement than was previously achieved (14.2%).

6.2.3 Summary of a measure to assess agreement in labour management.

A measure of the agreement between two sequences $A(t)$ and $B(t)$, normalised to the maximum agreement was derived. This measure considered the weighted and partial agreements between $A(t)$ and $B(t)$ according to an agreement matrix, $M_{A,B}$ and allowed partial temporal agreement for close temporal scoring using a window function $W(t-\tau)$. This measure of agreement is given by,

$$Agreement(A, B) = \frac{\max_{\tau=0}^{\tau=\pm T} \{w(\tau) \cdot \sum_{t=1}^N M_{a(t), b(t+\tau)}\} + \max_{\tau=0}^{\tau=\pm T} \{w(\tau) \cdot \sum_{t=1}^P M_{b(t), a(t+\tau)}\}}{\sum_{t=1}^N M_{a(t), a(t)} + \sum_{t=1}^P M_{b(t), b(t)}}$$

Where $M_{A(t), B(t)}$ was derived for the application of labour management and is given by,

		b(t)				
		1	2	3	4	5
M(A,B) =	a(t)	1	2	3	4	5
	2	1.5	2	1	0	0
	3	0	1	14	10	5
	4	0	0	10	25	20
	5	0	0	5	20	60

and where a cosine weighting function to measure temporal partial agreements was proposed.

$$W(\tau) = \cos \frac{\pi \tau}{8}$$

for $\tau = 0, \pm 1, \pm 2, \pm 3$

6.3 Derivation of the statistical distribution of agreement.

A method has been proposed to measure the agreement between any 2 given experts (or the system) for the management of labour. It could be that the proposed method is overly generous or perhaps even harsh which may bias the measured agreement towards high or low values. The question put another way is, what is the significance of a measured agreement? is a value of 80%, good? is a measure of 20% bad? Is the measure so biased that high agreement is inevitable?

These questions were addressed by considering the statistical parameters which described the distribution of the agreement measure. These parameters were,

1. The measure of agreement that could be expected by chance, $E[X]$ or μ .
2. The variation of the expected random chance agreements, $\text{Var}[X]$ or σ^2 .

With these parameters it would be possible to statistically test the significance of the measured agreements to investigate whether the experts and the system have performed worse than, no different from, or better than could have been expected by chance.

6.3.1 Derivation of the expected chance agreement.

To calculate the chance expected agreement at time t , it is necessary to assume that the length of sequences are infinite such that;

1. Edge effects are negligible. Edge effects occur at the beginning and end of the sequences because a partial agreement is given for the range of the weighting function, $t-3$ to $t+3$. When $t = 0, 1$ or 2 the number of adjacent scores which can be considered for a partial temporal agreement is reduced as scores before $t = 0$ do not exist. Similarly for the end of the sequence.
2. The validation protocol allowed the experts to conclude their review if they reached a score of 5 (caesarean section) or 3 in the second stage of labour (possible forceps delivery). For the derivation of the expected value it is assumed that sequences do not terminate when these conditions are met.

The *expected value* of a random variable, X, is the mean of the probability distribution and is written E[X] or μ . This is defined for a given probability distribution,

X	x(1)	x(2)	x(3)	x(n)
Probability	p(1)	p(2)	p(3)	p(n)

as,

$$\mu = E[X] = \sum_{i=1}^n p(i).x(i)$$

The expected value in the measurement of agreement is given by equation 6.14.

$$E[A, B] = \frac{E[Cac(A, B)] + E[Cac(B, A)]}{E[Aac(A)] + E[Aac(B)]} \quad (6.14)$$

With a symmetrical matrix and two sequences of equal length then, $E[Cac(A, B)] = E[Cac(B, A)]$ and $E[Aac(A)] = E[Aac(B)]$, this can be rewritten as equation 6.15.

$$E[A, B] = \frac{E[Cac(A, B)]}{E[Aac(A)]} \quad (6.15)$$

where, $E[Cac(A, B)] = E[Cac(B, A)] = E\left[\max_{\tau=0}^{\tau=\pm T} \{w(\tau) \cdot \sum_{t=1}^P M_{b(t), a(t+\tau)}\}\right]$

and, $E[Aac(A, B)] = E[Aac(B, A)] = E\left[\sum_{t=1}^P M_{a(t), a(t)}\right]$

and where the mapping matrix M(A,B) and weighting function W(t- τ) are,

		b(t)				
		1	2	3	4	5
M(A,B) =	a(t)	1	2	3	4	5
	2	1.5	2	1	0	0
	3	0	1	14	10	5
	4	0	0	10	25	20
	5	0	0	5	20	60

$$W(t - \tau) = \cos \frac{\pi \tau}{8}$$

for $\tau = 0, \pm 1, \pm 2, \pm 3$ which gives, $w(0) = 1.000$, $w(1) = 0.924$, $w(2) = 0.707$
and $w(3) = 0.383$

Calculation of the expected chance agreement, $E[Cac(B,A)]$.

It has been established that the experts recorded their scores with the overall probability distribution (to 3 d.p.);

Score (n)	1	2	3	4	5
Probability p(n)	0.435	0.438	0.071	0.039	0.017

Strategy.

Let $A(t)$ and $B(t)$ at time t , be represented by,

$A(t)$	$a(t-3)$	$a(t-2)$	$a(t-1)$	$a(t)$	$a(t+1)$	$a(t+2)$	$a(t+3)$
$B(t)$	$b(t-3)$	$b(t-2)$	$b(t-1)$	$b(t)$	$b(t+1)$	$b(t+2)$	$b(t+3)$

and the agreement measure given by,

$$\max \{w(t)M_{b(t)a(t)}, w(t-1)M_{b(t)a(t+1)}, w(t+1)M_{b(t)a(t-1)}, w(t-2)M_{b(t)a(t-2)}, \\ w(t+2)M_{b(t)a(t+2)}, w(t-3)M_{b(t)a(t-3)}, w(t+3)M_{b(t)a(t+3)}\}$$

Where $a(t)$ and $b(t)$ are scores, 1, 2, ... 5, and $w(t \pm \tau)$ are the cosine window coefficients. The maximum agreement is always when $b(t) = a(t)$. However, when $b(t) \neq a(t)$ the adjacent scores may provide a higher weighted agreement. The following algorithm was proposed to derive the chance expected agreement.

Step 1. assume a particular score for $b(t)$.

Step 2. for each $b(t)$, consider the possible scores for $a(t)$.

Step 3. for each $a(t)$, examine the possible adjacent scores which could achieve a higher agreement than, $w_0.M\{b(t), a(t)\}$.

repeat for each possible score of $b(t)$

Step 1. Assume $b(t) = 1$.

The probability that $b(t) = 1$ is $p(1)$. If this event happens then the ways of achieving agreement with $A(t)$ are shown in table 6.2,

A(t)	Agreement	Rank (r)
a(t) = 1	w(0).M(1,1) = 2	1
a(t) = 2	w(0).M(1,2) = 1.5	3
a(t±1) = 1	w(1).M(1,1) = 1.84	2
a(t±1) = 2	w(1).M(1,2) = 1.38	5
a(t±2) = 1	w(2).M(1,1) = 1.42	4
a(t±2) = 2	w(2).M(1,2) = 1.065	6
a(t±3) = 1	w(3).M(1,1) = 0.76	7
a(t±3) = 2	w(3).M(1,2) = 0.57	8
a(t±τ) = 3, 4, or 5	w(τ).0 = 0	9

Table 6.2: The possible agreements with A(t) when b(t) = 1.

These agreements were first ranked from the highest agreement measure (rank, r(1)) to the lowest, with ties given the same ranking.

Step 2. Consider each possible value of a(t),

2.1 Consider when a(t) = 1.

This has ranking = 1 and if it occurs will always be the maximum agreement. As such, the adjacent segments are not considered. The probability that this event will occur by chance is $p(b(t) = 1) * p(a(t) = 1)$. Thus, the expected chance agreement of this event is given by,

$$E[1,1,r(1)] = w(0).M(1,1).p(1).p(1)$$

2.2 Consider when a(t) = 2.

Here r(1), (a(t) = 1) cannot occur as it is assumed that a(t) = 2. From table 6.2 this event achieves an agreement ranked r(3) meaning it will be the maximum agreement so long as rank r(2) does not occur (a(t±1) ≠ 1). The probability that rank r(2) will not occur is $\text{prob}\{a(t±1) \neq 1\} = (1 - p(1))^2$. The expected chance agreement of this event is,

$$E[1,2,r(3)] = w(0).M(1,2).p(1).p(2).p(\overline{\text{rank } r(2)})$$

$$E[1,2,r(3)] = w(0).M(1,2).p(1).p(2).(1 - p(1))^2$$

Step 3. Consider higher agreements with adjacent scores.

3.1 a(t) = 2, rank r(2)

When a(t) = 2, rank r(2) is always the maximum agreement. This occurs when either a(t+1) = 1 or a(t-1) = 1 which has the probability $\{1 - (1 - p(1))^2\}$. The expected chance agreement is given by,

$$E[1,2,r(2)] = w(1).M(1,1).p(1).p(2).(1 - (1 - p(1))^2)$$

Aside.

To illustrate the derivation of $\{1 - (1 - p(x))^2\}$ consider a 6 sided dice, thrown three times. An analogous event is the probability of rolling a '6' on either the first, $t(1)$, or third, $t(3)$, throw which can be written $p\{t(1) = '6' \text{ or } t(3) = '6'\}$. This event can equally be expressed as $1 - \{\text{the probability of not throwing a '6' on the first throw AND not throwing a '6' on the third throw}\}$, which can be written,

$$\begin{aligned} & \{1 - (p(t(1) \neq '6' \text{ and } t(3) \neq '6'))\} \\ & = \{1 - (1 - p('6')) * (1 - p('6'))\} \\ & = \{1 - (1 - p('6'))^2\} \end{aligned}$$

or for the general case this is written $\{1 - (1 - p(x))^2\}$

It should be noted that this probability is *not* $p('6') + p('6') = 2.p('6')$.

2.3 Consider when $a(t) = 3$.

This event achieves 0 agreement and is ranked $r(9)$. If this event occurs, then $a(t) \neq 1$ and $a(t) \neq 2$, thus ranks $r(1)$ and $r(3)$ cannot occur. There are 6 other events where the scores adjacent to $a(t)$ would be $> r(9)$. These are, $r(2)$, $r(4)$, $r(5)$, $r(6)$, $r(7)$ and $r(8)$.

3.1 $a(t) = 3$, rank $r(2)$.

If $a(t) = 3$, then rank $r(2)$ is the maximum agreement possible and occurs when either, $a(t+1) = 1$ or $a(t-1) = 1$, which has the probability $\{1 - (1 - p(1))^2\}$. The expected chance agreement is given by,

$$E[1,3,r(2)] = w(1).M(1,1).p(1).p(3).(1 - (1 - p(1))^2)$$

3.2 $a(t) = 3$, rank $r(4)$.

This event will be the maximum if rank $r(2)$ does not occur. The expected chance agreement is,

$$\begin{aligned} E[1,3,r(4)] &= w(2).M(1,1).p(1).p(3).(1 - (1 - p(1))^2).p(\overline{\text{rank } r(2)}) \\ E[1,3,r(4)] &= w(2).M(1,1).p(1).p(3).(1 - (1 - p(1))^2).(1 - p(1))^2 \end{aligned}$$

3.3 $a(t) = 3$, rank $r(5)$.

This event will be the maximum if ranks $r(2)$ and $r(4)$ do not occur. The expected chance agreement is,

$$\begin{aligned} E[1,3,r(5)] &= w(1).M(1,2).p(1).p(3).(1 - (1 - p(2))^2).p(\overline{\text{rank } r(2)}).p(\overline{\text{rank } r(4)}) \\ E[1,3,r(4)] &= w(2).M(1,1).p(1).p(3).(1 - (1 - p(1))^2).(1 - p(1))^2.(1 - p(1))^2 \end{aligned}$$

The algorithm then continues

Step 3 considers the expected agreements for ranks $r(6)$, $r(7)$ and $r(8)$.

Then, Step 2 is continued by considering when $a(t) = 4$ and $a(t) = 5$. Step 3 is carried out for each score.

Then, back to step 1 which next assumes $b(t) = 2$ and so on.

This derivation was too lengthy to obtain by hand and so the algorithm was implemented in software and is given in appendix I. This algorithm gives the total expected chance agreement of $b(t)$ with $a(t-3) \dots a(t+3)$, when $b(t) = 1$, as,

$$E[\text{Cac}(1,A)] = E[1,1 r(1)] + E[1,2 r(2)] + E[1,2 r(3)] + E[1,3 r(2)] + E[1,3 r(4)] + E[1,3 r(5)] + E[1,3 r(6)] + E[1,3 r(7)] + E[1,3 r(8)] + \dots + E[1,5 r(8)]$$

$$\underline{E[\text{Cac}(1,A)] = 0.802745}$$

and the total expected chance agreement of $\text{Cac}(B,A)$ as,

$$E[\text{Cac}(B,A)] = E[\text{Cac}(1,A)] + E[\text{Cac}(2,A)] + E[\text{Cac}(3,A)] + E[\text{Cac}(4,A)] + E[\text{Cac}(5,A)]$$

$$E[\text{Cac}(B,A)] = 0.802745 + 0.809399 + 0.419290 + 0.309097 + 0.152503$$

$$\underline{E[\text{Cac}(B,A)] = 2.493}$$

The expected chance auto-agreement coefficient $Aac(A) = Aac(B)$ is,

$$Aac(B) = Aac(A) = \sum_{a(t)=1}^5 M\{a(t), a(t)\} \cdot p(a(t))$$

$$\underline{Aac(A) = Aac(B) = 4.728}$$

Which gives the normalised chance expected agreement as,

$$E[A, B] = \frac{\text{Cac}(A, B)}{Aac(A)} = \frac{2.493}{4.728}$$

$$\underline{E[A, B] = 0.527 = 52.7\%}$$

This derived figure was checked by measuring the chance agreement between 2 random sequences, using a random number generator with a probability distribution of scores (1,2, .. 5) equal to the experts. Each sequence was of length, $N = 1000$, and 30,000 pairs were generated for which the agreements were measured. The random number generator had a periodicity of 2^{32} . The chance agreement expressed as a percentage was plotted against the frequency distribution normalised to f_{\max} and is shown in figure 6.1. The recorded frequencies outside this range were all zero.

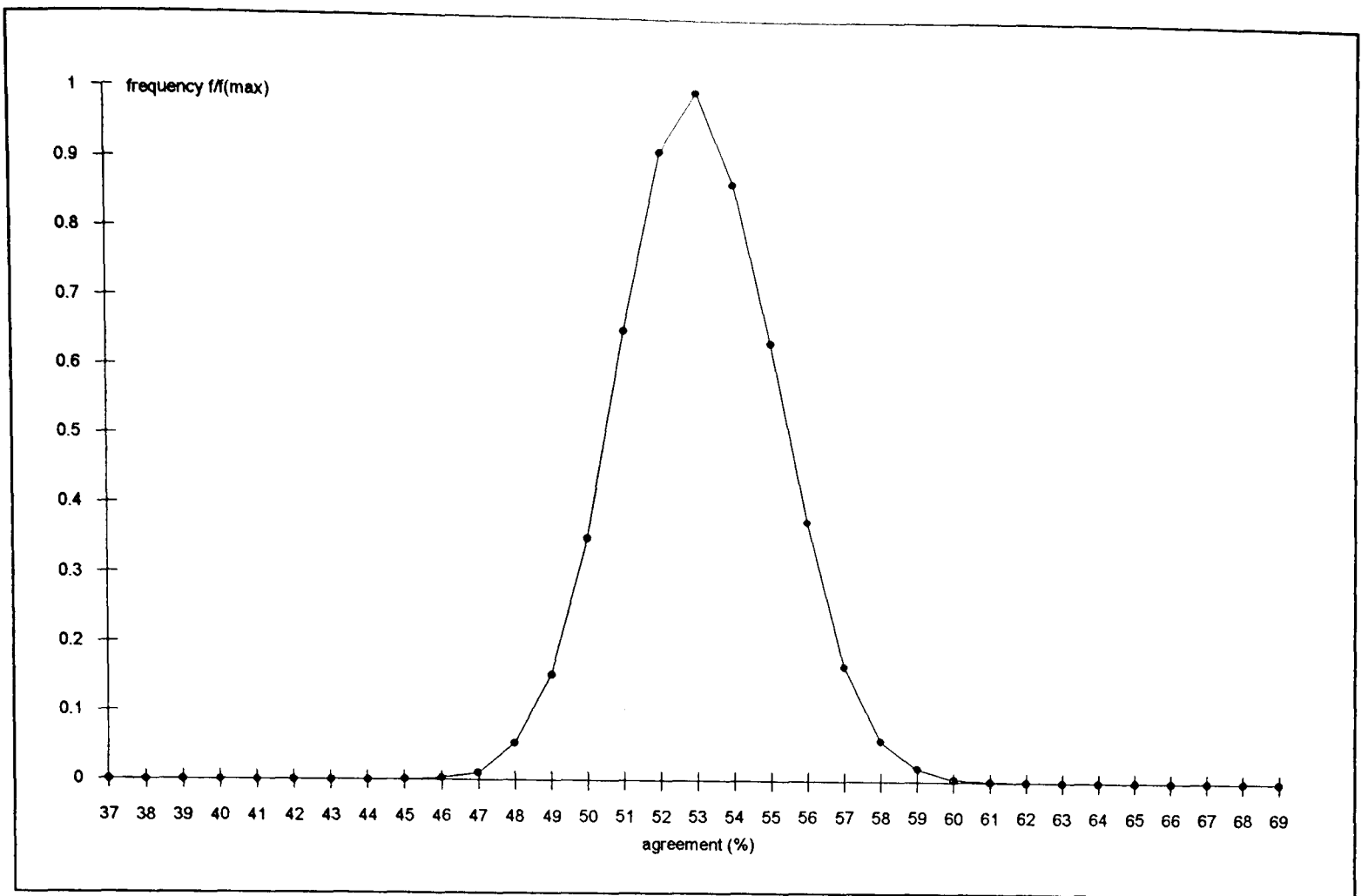


Figure 6.1: The distribution of random agreements.

This distribution is approximately normally distributed with mean 53.0%.

This confirms the approach for the derivation of the expected agreement, μ , but it will be recalled that the derivation relied on two simplifications, which briefly restated are; edge effects were negligible and sequences never terminated prematurely. The validity of these simplifications were investigated by considering the agreement which could be obtained from plausible random numbers.

Two sets of plausible random numbers which obeyed the study protocol were generated for each of the 50 cases and are shown on the score tables in appendix H as R1 and R2. The distribution of generated scores were confirmed to have the same probability distribution as the experts scores. The average agreement over the 50 cases was calculated to be 50.9% which is 3.4% lower than that derived with simplifications. This confirms that the simplifications do not overly effect the derived measure.

6.3.2 Derivation of the expected variance, σ^2 .

The variance of a probability distribution is given by,

$$\text{variance} = \sum p(i).(x - \mu)^2$$

The variance $\text{Var}(X)$ has the symbol σ^2 and is the expected value of the square of the deviation from the mean, $E[X - \mu]^2$. This can be simplified, $\sigma^2 = E[X^2] - \mu^2$

The expected value, μ , has been derived. The value of $E[X^2]$ was derived with a similar strategy to that used for $E[X]$. Where now,

$$E[(A,B)^2] = \frac{E[\text{Cac}(A,B)^2]}{E[\text{Aac}(A)^2]}$$

The algorithm in appendix I was modified to obtain this value which was found to be,

$$E[(A,B)^2] = 0.502 \text{ or } 50.2\%$$

This gives,

$$\sigma^2 = 0.502 - 0.527^2$$

$$\underline{\underline{\sigma^2 = 0.224}}$$

6.4 Statistical tests to investigate agreement.

It was proposed to examine the nature of the agreement results obtained by the experts and the system using two statistical tests.

1. The kappa coefficient of inter-rater agreement.

This measure compares inter-rater agreements and examines the significance of measured agreement in excess of that which could have been expected by chance.

2. The Kruskal-Wallis one-way analysis of variance by ranks.

This non-parametric statistic examines the characteristics of more than 2 distributions to establish whether any differences in the distributions represent genuine population differences or whether the variations are those to be expected from random samples of similar populations.

6.4.1 The kappa coefficient measure of inter-rater agreement.

The kappa statistic was first proposed by Cohen (1960) and has been adapted to allow partial scoring agreements (Cohen, 1968). Much work in this area has been done by Landis and Koch (1977) and an informative text is given by Fleiss (1981). The kappa statistic is given by equation 6.16.

$$kappa = \frac{o - e}{1 - e} \quad (6.16)$$

Where, o , is the observed or measured agreement and, e , is the level of agreement the raters could have achieved by chance. The numerator, $o - e$, is therefore the observed agreement in excess of that which could have been expected by chance and the denominator, $1 - e$, is the maximum agreement the raters could have achieved in excess of random chance. The kappa statistic has desirable features; it reaches a maximum of 1.0 for perfect agreements, scores 0 when the agreement is equal to that expected by chance, and is less than 0, when the agreement is less than that expected by chance. The exact lower limit depends on the given application, but if the measured agreement takes the range 0 to 1, where 0 indicates no agreement, then the minimum value of kappa is $-e / (1 - e)$ which is -1 when $e = 0.5$. It has been possible to derive the expected value, $E[o] = \mu_o = e$, for the agreement measures and so the proposed form of the kappa equation becomes equation 6.17.

$$kappa = \frac{o - \mu_o}{1 - \mu_o} \quad (6.17)$$

It has been suggested (Donker, 1991; Fleiss, 1981) that ranges have been characterised from which kappa values can be interpreted; it was stated that kappa values "larger than 0.75 may be interpreted to represent excellent agreement beyond that which could be achieved by chance alone, values below 0.40 may be interpreted to represent poor agreement beyond chance, and values between 0.40 and 0.75 may be interpreted to represent fair to good agreement beyond chance." This classification was suggested by both authors to reflect the original ranges proposed by Landis and Koch (1977). In fact this is misleading. The original classification of kappa made Landis and Koch was,

Kappa statistic	Strength of Agreement
< 0.00	Poor
0.00 - 0.20	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 1.00	Almost perfect

There are clearly differences between these interpretations, particularly in the low kappa scores. This possibly indicates the dangers of adopting linguistic variables to describe data.

In fact, Landis and Koch did not intend for their classification to be taken as universal, as implied by Donker (1991). In their original article the authors state that, "these divisions are clearly arbitrary", and were intended to, "provide useful benchmarks" for the discussion of a specific example presented in their paper. This example considered 2 sets of data, where the

number of samples was, $n = 149$ and 69 . It may be that the proposed benchmarks were applicable for their example but it does not seem reasonable that benchmarks could be considered as universal indicators. The significance of any statistic, kappa included, will depend on the population distribution from which it was sampled and most importantly, the number of classifications or scores considered. For example, an average kappa of 0.8 measured for $n = 3$ does not in any way carry the same significance as a kappa of 0.8 when $n = 1000$. The preferred approach would be to test the significance of a measure using conventional statistical sampling theory. This approach will now be developed formally.

In simple random sampling, if $X(1), X(2) \dots X(n)$ are n random samples observed from a population distribution whose mean is, μ , and variance is, σ^2 , then $X(1), X(2) \dots X(n)$, are independent random variables, each of which has the same distribution as the parent population. If \bar{X} is the mean of such a sample of size n , then it can be shown that,

$$\underline{\underline{E[\bar{X}] = \mu}}$$

$$\underline{\underline{Var[\bar{X}] = \frac{\sigma^2}{n}}}$$

The standard deviation of the sampling distribution of means is called the standard error of the mean and is given by,

$$\underline{\underline{standard\ error\ of\ the\ mean\ s.e. = \frac{\sigma}{\sqrt{n}}}}$$

The mean, μ_0 , and variance, σ_0^2 , of the chance agreements have been found and may be used to derive, the mean μ_k , and variance, σ_k^2 of the distribution of kappa. These kappa parameters could then be used to test various hypotheses regarding a measured mean kappa agreement, \bar{k} , in the usual way by referring the quantity,

$$Z = \frac{\bar{K} - \mu_k}{\frac{\sigma_k}{\sqrt{n}}} = \frac{\bar{K}}{s.e.(\bar{k})} \quad (as\ \mu_k = 0) \quad (6.18)$$

to the standard Normal tables. Where, n , is the number of segments and $s.e.(\bar{k})$ is the standard error of the means of kappa.

Derivation of the expected kappa value, μ_k and expected variance, σ_k^2 .

In order to calculate the mean, μ_k , and variance, σ_k^2 we must consider the transformation of μ_0 and σ_0^2 represented in equation 6.17 as,

$$k = \frac{o - \mu_o}{1 - \mu_o}$$

In statistical expectation, the general linear transformation of a random variable X, by the constants, a and b is given by.

$$E[bX + a] = bE[X] + a = b\mu + a \quad (6.19a)$$

$$Var[bX + a] = b^2 Var[X] = b^2 \sigma^2 \quad (6.19b)$$

Kappa is given by, $k = \frac{o - \mu_o}{1 - \mu_o}$

For the mean, μ_k ,

$$\mu_k = E[K] = \frac{1}{1 - \mu_o} E[o - \mu_o]$$

$$= \frac{1}{1 - \mu_o} \cdot \{E[o] - \mu_o\}$$

$$\underline{\underline{\mu_k = E[K] = 0}} \quad (\text{since } E[o] = \mu_o)$$

For the variance, σ_k^2

$$\sigma_k^2 = Var[K] = Var\left[\frac{o - \mu_o}{1 - \mu_o}\right] = Var\left[\frac{o}{1 - \mu_o} - \frac{\mu_o}{1 - \mu_o}\right]$$

which is in the form $var[bX + a]$ and using (6.19b) gives,

$$\sigma_k^2 = Var[K] = \frac{\sigma_o^2}{(1 - \mu_o)^2}$$

and substituting for σ_o^2 and μ_o gives,

$$\underline{\underline{\sigma_k^2 = \frac{0.224}{(1 - 0.527)^2} = 0.998}}$$

6.4.2 The Kruskal-Wallis one-way analysis of variance by ranks.

The Kruskal-Wallis one-way analysis of variance by ranks is a test for deciding whether c independent samples are from different populations (Kruskal, 1952; Kruskal and Wallis, 1952). This is a popular statistical test and has been described in general texts (Lehmann, 1975; Hettmansperger, 1984). This test determines whether the variations between random samples are no more than one would expect from samples obtained from the same population, or whether the variations are large enough to signify that the samples are in fact from different populations. If θ_j is the median of the j th sample, then this statistic tests the null hypothesis that the sample medians are all equal $\theta_1 = \theta_2 = \dots = \theta_c$ against the alternative hypothesis that at least two medians are different $\theta_i \neq \theta_j$ for some samples i and j . The samples are first cast into a table where each column represents each sample.

Group				
1	2	•	•	c
X(1,1)	X(1,2)	•	•	X(1,c)
X(2,1)	X(2,2)	•	•	X(2,c)
•	•	•	•	•
X(n(1),1)	•			X(n(c),c)
	X(n(2),2)			

Where X_{ij} is the datum for the i th observation in the j th group and, $n(j)$, is the number of observations in the j th group.

The next step is to replace each of the N observations, $N = n(1) + n(2) + \dots + n(c)$, with their overall ranking. That is, all of the observations from the c samples are ranked as a single series. The smallest score is replaced by rank 1, the next by rank 2 and the largest by rank N . The sum of the ranks is then found for each column from which the column average can be found. This test makes use of the fact that if the samples were drawn from identical populations, the average rank of each column should be similar, whereas, if the samples were drawn from different distributions, with different medians, then the average ranks would differ. The Kruskal-Wallis statistic (KW) is computed using equation 6.20.

$$KW = \frac{12}{N(N+1)} \sum_{j=1}^c n(j)(\bar{R}_j - \bar{R})^2 \quad (6.20)$$

Where c = number of samples or groups,
 $n(j)$ = the number of observations in the j th group,
 N = the total number of observations,
 R_j = sum of the ranks in the j th group,
 \bar{R}_j = average of the ranks of the j th group,
 \bar{R} = average of the ranks for the combined samples.

When there are more than $c = 3$ groups, and when the number of observations in each group, $n_j > 5$, then sampling distribution of KW approximates the chi-squared, χ^2 , distribution with degrees of freedom, $df = c - 1$.

When ties occur between two or more scores, each score is given the mean of the ranks for which they tie. As ties influence the KW statistic, it is corrected by dividing equation 6.20 by equation 6.21.

$$1 - \frac{\sum_{i=1}^g (t_i^3 - t_i)}{N^3 - N} \quad (6.21)$$

where g = number of groupings of different tied ranks,
 t_i = number of tied ranks in the i th grouping.

The corrective effect is often negligible when no more than 25% of observations tie.

If, KW, reaches significance, it indicates that at least one of the groups is different from at least one of the others. It does not indicate which ones are different, nor does it indicate how many are different. This information can subsequently be found by considering the null hypothesis that, the medians of any 2 given groups, u and v , are the same, $\theta_u = \theta_v$, against the alternative hypothesis, that the medians are different, $\theta_u \neq \theta_v$. This is achieved by obtaining the differences of the mean of the ranks, $|\bar{R}_u - \bar{R}_v|$, for all groups. When the sample size is large, these differences are approximately normally distributed. However, since there are a large number of differences, and because the differences are not independent, the comparison technique requires adjustment. If KW reaches significance at the, α level, then the significance of individual pairs of differences are tested using the inequality of equation 6.22.

$$|\bar{R}_u - \bar{R}_v| \geq z_{\alpha/c(c-1)} \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_u} + \frac{1}{n_v} \right)} \quad (6.22)$$

Where, $z_{\alpha/c(c-1)}$ is the value from the unit normal distribution above which lies $\alpha/c(c - 1)$.

6.5 Application of statistical hypothesis tests.

6.5.1 Did the experts or system obtain an average agreement in any case significantly better than, or worse than expected by chance?

For each case, the mean agreement the system scored with the experts, $o(\text{system})$, and the mean agreement the experts scored with each other, $o(\text{experts})$, was obtained. The case length, n , was taken as the total number of 15 minute segments in each case. The kappa values, $k(\text{system})$ and $k(\text{expert})$, were then obtained with equation 6.17. These kappa values were used to test the null hypothesis that neither the experts, nor the system, obtained agreements significantly different from those expected by chance and two alternative hypothesis were considered.

1. If the agreement reached significance and was less than expected by chance, then it was concluded that there was significant disagreement.
2. If the agreement reached significance and was greater than expected by chance then it concluded that there was significant agreement.

Although two alternative hypothesis have been formed, the test was one-tailed because each predicts the direction of the agreement. In this test, the region of rejection should not be too small to prevent a measure of agreement from reaching significance (in either direction) when the number segments, n , scored in a case was small. Therefore the significance level was chosen as, $\alpha = 0.1$. It has been found that the sampling distribution of the means of kappa is normally distributed with $\mu_k = 0$ and $\sigma_k^2 = 0.998$. This allows us to refer equation 6.18.

$$z = \frac{\bar{k}}{\sigma_k / \sqrt{n}} \quad (6.18)$$

to the standard normal tables. The results obtained are shown in table 6.3.

Case	Segments (n)	Measured agreement		$\overline{\text{kappa}}$		Z values	
		o (experts)	o (system)	k (experts)	k (system)	z (experts)	z (system)
1	12	68.1	71.5	0.33	0.40	1.15	1.39
2	45	43.7	44.7	-0.19	-0.17	-1.28	-1.14
3	47	52.2	54.8	-0.01	0.05	-0.07	0.34
4	20	57.9	57.4	0.11	0.10	0.49	0.45
5	26	58.7	63.4	0.13	0.23	0.66	1.18
6	10	92.1	91.6	0.83	0.82	2.63	2.60
7	28	67.8	76.6	0.32	0.50	1.70	2.65
8	16	76.5	63.6	0.50	0.23	2.00	0.92
9	7	80.5	76.2	0.59	0.50	1.56	1.33
10	9	64.4	70.2	0.25	0.37	0.75	1.11
11	16	80.6	65.2	0.59	0.26	2.36	1.04
12	19	55.7	62.4	0.06	0.21	0.26	0.92
13	56	59.4	65.1	0.14	0.26	1.05	1.95
14	16	65.0	61.7	0.26	0.19	1.04	0.76
15	20	66.3	67.9	0.29	0.32	1.30	1.43
16	47	44.0	52.1	-0.18	-0.01	-1.24	-0.07
17	20	57.7	46.7	0.10	-0.13	0.45	-0.58
18	36	66.6	66.6	0.29	0.29	1.74	1.74
19	15	59.1	57.2	0.13	0.10	0.50	0.39
20	23	75.0	81.2	0.47	0.60	2.26	2.88
21	16	90.6	92.0	0.80	0.83	3.21	3.33
22	20	66.3	67.9	0.29	0.32	1.30	1.43
23	52	70.0	73.3	0.37	0.44	2.67	3.18
24	11	64.0	52.4	0.24	-0.01	0.80	-0.03
25	14	80.4	76.8	0.58	0.51	2.17	1.91
26	31	75.9	76.9	0.49	0.51	2.73	2.85
27	8	79.1	75.9	0.56	0.49	1.59	1.39
28	26	69.5	73.0	0.36	0.43	1.84	2.20
29	52	84.5	79.0	0.67	0.56	4.84	4.05
30	11	63.9	49.4	0.24	-0.07	0.80	-0.23
31	16	97.1	97.6	0.94	0.95	3.77	3.81
32	17	49.8	40.2	-0.06	-0.26	-0.25	-1.07
33	15	76.2	75.3	0.50	0.48	1.94	1.86
34	23	74.9	78.7	0.47	0.55	2.26	2.64
35	9	81.5	72.8	0.61	0.43	1.83	1.29
36	38	63.7	64.9	0.23	0.26	1.42	1.61
37	59	65.8	63.4	0.28	0.23	2.16	1.77
38	14	81.3	29.1	0.60	-0.50	2.25	-1.87
39	27	90.8	91.9	0.80	0.83	4.17	4.32
40	11	57.7	38.7	0.11	-0.30	0.37	-1.0
41	15	63.6	61.7	0.23	0.19	0.89	0.74
42	17	61.7	54.5	0.19	0.04	0.78	0.17
43	19	83.1	76.2	0.64	0.50	2.80	2.18
44	20	64.8	31.8	0.26	-0.44	1.17	-1.97
45	14	68.7	73.0	0.34	0.43	1.27	1.61
46	34	78.4	81.8	0.54	0.61	3.16	3.56
47	20	92.5	92.4	0.84	0.84	3.76	3.76
48	38	65.0	46.0	0.26	-0.14	1.61	-0.86
49	29	89.2	92.3	0.77	0.84	4.15	4.53
50	28	88.0	91.3	0.75	0.82	3.98	4.35

Table 6.3: Statistical significance in the measures of agreement for each case.

The average agreement the system scored with the experts, o(system), and the average agreement the experts scored with each other, o(experts), was calculated for each case. These were expressed as kappa values, k(system) and k(experts), which were referred to the standard normal tables to obtain, z(system) and z(experts) to test the level of agreement above chance.

At the $\alpha = 0.1$ level of significance the critical value of, z , from the standard normal tables is, $z = 1.27$. If for a given case, $z \leq -1.27$, then it was concluded that the average agreement was significantly less than expected by chance and represented disagreement. If, $z \geq 1.27$, then the average agreement was significantly better than expected by chance and if, $-1.27 \leq z \leq 1.27$, then the average measured agreement was not significantly different from that expected by chance. So from table 6.3 it can be found that,

1. The average agreement the experts scored with each other was significantly better than was expected by chance in 31 cases.
2. The average agreement the system scored with the experts was significantly better than was expected by chance in 29 cases.
3. The experts disagreed with each other in the management of 1 case (case 2).
4. The system disagreed with the experts in the management of 2 cases (38 and 44).

6.5.2 Did each expert and the system manage the cases similarly?

Unlike the previous test which considered average agreements, this test considered the agreement measures of the experts and the system obtained on an individual basis.

If an expert had managed a case differently from the other experts, then they would have obtained lower measures of agreement. If the system did not embody expertise then it would have obtained lower measures of agreement with the experts than they obtained with each other.

For each case, each review (A1, A2, B1, ..., S2) was compared with all 36 reviews (not including the random numbers). However, 2 comparisons for each review need not be considered here; the first is obviously the reviews agreement with itself (e.g. A1 with A1 = 100%) and the second is the inter-reviewer agreement (e.g. A1 with A2 = 69%) which will be considered later. Therefore, each of the 36 reviews has a set of 34 measures of agreement with other reviewers. These are the rows and columns in each agreement table (chapter 5, section 5.5.2, table 5.8 and appendix H). The Kruskal-Wallis test was used to establish if all sets for a given case contained similar levels of agreement. If this test reached significance then it indicated that at least 2 sets were significantly different. Here, the set with the higher agreement would represent the management preferred by the other reviewers and the lower set would represent the management least preferred. A measure of the performance of each reviewer could be obtained by considering the number of cases for which they obtained a significantly lower set of agreements than at least one other reviewer.

For each case considered separately, the null hypothesis was that; there was no difference in the levels of agreements that each reviewer obtained. The alternative hypothesis was that at least two reviews differed in the levels of agreement they obtained. The total number of agreements per case, $N = 36 \times 34 = 1224$. The significance level was chosen to be, $\alpha = 0.05$. Since the sample sizes exceed 5, the sampling distribution of the Kruskal-Wallis statistic, KW, approximates the chi-squared, χ^2 , distribution with degrees of freedom, $df = c - 1 = 35$ for the number of groups, c . This significance level and degrees of freedom gives a critical value for $KW = 50.96$.

It was found that the KW statistic reached significance in all 50 cases which implied that in each case there were at least 2 reviewers with different levels of agreement measures.

The next stage was to investigate for each reviewer in each case, whether any other reviewer obtained a significantly higher level of agreement. This was achieved by rearranging equation 6.22 to obtain equation 6.23 to specifically test whether a given reviewer, u , has significantly lower agreements than another reviewer, v .

$$\bar{R}_u \geq \bar{R}_v + z_{\alpha/c(c-1)} \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_u} + \frac{1}{n_v} \right)} \quad (6.23)$$

for each reviewer u , and every combination of other reviews v , where,

\bar{R}_u = average of the ranks of the u th group,

\bar{R}_v = average of the ranks of the v th group,

$\alpha/c(c - 1) = 0.05/36(36 - 1) = 0.00004$, which gives,

$z_{0.00004} = 3.95$ from the standard normal tables.

$n_u = n_v =$ the number of measures in each group = 34

$N = 36 \times 34 = 1224$.

Substituting these values reduces the inequality to,

$$\bar{R}_u \geq \bar{R}_v + 338.64 \quad (6.24)$$

For each reviewer, the cases were found where at least one other review had obtained a significantly higher level of agreement. These were represented into tables which are given in appendix J. The results obtained for expert A are shown in table 6.4.

- The cases where expert A obtained a significantly lower agreement for his recommended management than at least one other reviewer.
- The median agreement the reviewer obtained in the specific case is shown as θ .

- The reviews which obtained a significantly higher agreement are identified.
- The total number of reviews which obtained significantly higher agreements for the specific case is shown.

case	review	θ	Reviews which have significantly higher agreement	total
4	A1	46	E1 E2 G1 I2 J2 K1 L1 O1 O2	9
9	A2	73	A1 B2 C1 F1 F2 G1 I1 I2 K2 L1 L2 M2 N2 O1 O2 P2	16
11	A1	68	A2 B2 C1 C2 D1 E1 E2 F2 G1 H1 I1 I2 K1 L1 M2 N2	16
20	A1	64	C2 D1 D2 E1 E2 H1 H2 K1 N1 N2 O2 P2 S1 S2	14
21	A1	68	A2 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H1 H2 I1 I2 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2 S1 S2	32
23	A1	75	A2	1
25	A2	72	B2 D1 D2 E1 E2 G1 H2 J1 J2 K1 L1 L2 M1 N1 N2	15
27	A2	52	A1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 H1 H2 I2 J2 K1 K2 L1 L2 M2 N1 O1 O2 P2 Q1	25
28	A1	63	B2 D1 E1 E2 F2 G1 G2 I1 J2 K2 L2 M2 N2 O1 O2 P2 Q1 S1 S2	19
29	A2	83	D2 L2	2
30	A1	42	A2 D1 D2 E1 H2 J2 L1 L2 M1 M2 O2 P1	12
34	A1	56	A2 C2 D1 D2 E1 E2 G1 H2 I1 I2 K2 L1 L2 M1 M2 O2 P2 S1 S2	19
35	A1	70	A2 D1 D2 E1 E2 G1 H1 I2 J1 J2 K2 L1 N2 O2	14
37	A1	50	C2 D2 H2 L1 L2 O1	6
38	A1	70	C1 C2 D1 D2 E1 E2 I1 I2 N2	9
39	A1	89	B1 B2 C1 D1 D2 E1 E2 F1 F2 G1 G2 H2 I1 I2 J2 K1 K2 L1 L2 M1 M2 N2 O1 P1 P2 S1 S2	27
39	A2	85	B1 B2 C1 D1 D2 E1 E2 H2 I1 I2 J2 K1 L1 L2 O1 P1 P2 S1 S2	19
50	A2	64	A1 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H2 I1 I2 J1 J2 K1 K2 M1 M2 N1 O2 Q1 S1 S2	27

Table 6.4: Cases where expert A obtained lower agreement than at least one other reviewer.

It can be seen from table 6.4, that the majority of other reviews (> 17) obtained significantly higher agreements than either expert A's first review, A1, or second review, A2, in 6 cases (21, 27, 28, 34, 39, 50) and 7 reviews. The results obtained for all experts and the system are expressed in this way in table 6.5.

Reviewer	Number Cases	Number Reviews
A	6	7
B	16	22
C	6	7
D	3	3
E	1	1
F	5	6
G	2	2
H	9	13
I	1	1
J	11	11
K	2	2
L	3	3
M	7	9
N	4	4
O	4	4
P	5	6
Q	28	36
System	5	10

Table 6.5 The number of cases and reviews where each reviewer obtained significantly lower agreements than the majority.

It was found that;

1. The system obtained significantly lower agreement than the majority of reviewers in 5 cases of the 50.
2. Most experts obtained significantly lower agreement than the majority of reviewers in 5 cases or less.
3. Experts B, H, J and Q, obtained significantly lower agreements than the majority of reviewers most often.
4. Expert Q deserves special mention here because in over half the cases, he obtained significantly lower agreements than the majority.

This test, by itself may not separate the good performers from the poor performers, although it contributes valuable evidence towards this. It may be that for some cases, those with lower agreements have managed the labours better than the rest.

The results obtained for the system are encouraging because in 45 out of 50 cases, the level of agreement that the system achieved with the experts was not significantly lower than the experts achieved with each other.

6.5.3 Assessment of each reviewers overall inter- and intra-agreement.

Each reviewers average agreement with the other experts (inter-agreement) and consistency (intra-agreement) was calculated. These measures were used to determine whether each expert and the system was able to agree and be consistent in excess of that expected by chance.

The agreement the system obtained with the experts for each of its reviews, S1 and S2, was averaged over all 50 cases. Similarly for each expert review, the average agreement with the remaining experts was found. The average of all inter-expert agreements was obtained to give a single measure of the overall inter-expert agreement, \bar{Ex} . In addition, the agreement that the two sets of plausible random numbers obtained with themselves and with the experts was found. The average consistency of each reviewer and the plausible random numbers was taken as the agreement measured between their 2 reviews of each case, averaged over the 50 cases. These results were then expressed as kappa values.

These kappa values were used to test the null hypothesis that no expert, nor the system, obtained an average agreement significantly different from that expected by chance. Two alternative hypothesis were considered;

1. If a reviewer's average agreement reached significance and was less than chance then it was concluded that the reviewer had significantly disagreed with the other experts in the management of the 50 cases.
2. If a reviewer's average agreement reached significance and was greater than expected by chance then it concluded that the reviewer had significantly agreed with the management of the other experts in the 50 cases.

Although two alternative hypothesis have been formed, the test was one-tailed because each predicts the direction of the outcome. The total number of segments, n , the experts scored was large, $n = 1129$. The level of significance was chosen to be, $\alpha = 0.01$. The sampling distribution of the means of kappa has been found to be normally distributed with $\mu_k = 0$ and $\sigma_k^2 = 0.998$. This allows equation 6.18,

$$z = \frac{\bar{k}}{\sigma_k / \sqrt{n}} \quad (6.18)$$

to be referred to the standard normal tables. The inter-agreement results are presented in table 6.6, the intra-agreement results are shown in table 6.7.

Inter-agreement			
Reviewer	\bar{O}_{inter}	\bar{K}_{inter}	z
A1	69.61	0.36	12.12
A2	73.06	0.43	14.48
B1	60.67	0.17	5.72
B2	65.18	0.26	8.75
C1	71.94	0.41	13.80
C2	72.71	0.42	14.14
D1	73.13	0.43	14.48
D2	73.61	0.44	14.81
E1	73.76	0.45	15.15
E2	73.35	0.44	14.81
F1	66.87	0.30	10.10
F2	70.92	0.39	13.13
G1	73.40	0.44	14.81
G2	70.66	0.38	12.79
H1	64.23	0.24	8.08
H2	69.76	0.36	12.12
I1	73.85	0.45	15.15
I2	74.69	0.46	15.49
J1	68.93	0.34	11.45
J2	72.32	0.41	13.80
K1	73.60	0.44	14.81
K2	72.04	0.41	13.80
L1	73.92	0.45	15.15
L2	73.37	0.44	14.81
M1	66.02	0.28	9.43
M2	70.05	0.37	12.46
N1	70.94	0.39	13.13
N2	71.48	0.39	13.13
O1	69.32	0.35	11.78
O2	74.10	0.45	15.15
P1	70.01	0.37	12.46
P2	72.60	0.42	14.14
Q1	59.54	0.14	4.71
Q2	56.80	0.09	3.03
S1	67.14	0.30	10.10
S2	67.51	0.31	10.44
\bar{Ex}	70.19	0.37	12.46
\bar{Ran}	47.1	-0.12	-4.04

Table 6.6: Each reviewers average inter-agreement.

The average inter-agreement was calculated for the experts, the system and the plausible random numbers, \bar{O}_{inter} . This was expressed as a kappa value, \bar{K}_{inter} from which the z value from the standard normal tables was obtained.

Intra-agreement			
Reviewer	\bar{O}_{intra}	\bar{K}_{intra}	z
A	76.42	0.50	16.83
B	78.68	0.55	18.52
C	80.00	0.58	19.53
D	82.32	0.63	21.21
E	82.10	0.62	20.87
F	81.20	0.60	20.20
G	77.28	0.52	17.50
H	78.62	0.55	18.52
I	89.04	0.77	25.92
J	73.18	0.43	14.48
K	81.00	0.60	20.20
L	83.16	0.64	21.55
M	84.18	0.67	22.56
N	82.56	0.63	21.21
O	78.44	0.54	18.18
P	81.58	0.61	20.54
Q	74.16	0.45	15.15
S	99.16	0.98	32.99
\bar{Ex}	80.23	0.58	19.53
\bar{Ran}	50.24	-0.05	-1.68

Table 6.7: Each reviewer's average intra-agreement (consistency).

The average intra-agreement was calculated for the experts, the system and the plausible random numbers, \bar{O}_{intra} . This was expressed as a kappa value, \bar{K}_{intra} from which the z value from the standard normal tables was obtained..

At the $\alpha = 0.01$ level of significance the critical value of z from the standard normal tables is $z = 2.32$. If a reviewer obtained $z \leq -2.32$ then it was concluded that, on average, the reviewer disagreed with the other experts. If $z \geq 2.32$ then the reviewer's average agreement was significantly better than expected by chance and if $-2.32 \leq z \leq 2.32$ then the average measured agreement was not significantly different from that expected by chance. From tables 6.6 and 6.7 it can be found that;

Inter-agreement results.

1. All experts and the system obtained an average agreement significantly better than expected by chance. For the most part (with the exception of Q1 and Q2), these results were highly significant; a z value of > 5.0 for instance has an associated probability, $p < 0.0000003$.
2. The system's average inter-agreement was within the range of the experts' agreement.

3. Expert Q obtained much the lowest inter-agreement. For his second review, this was only just significantly above that expected by chance.
4. The plausible random numbers disagreed with the experts.

Intra-agreement.

5. The system and the experts obtained a consistency significantly better than could have been expected by chance. Again, these results were highly significant.
6. The system achieved almost perfect consistency ($\bar{O}_{intra} = 99.16\%$, $\bar{k}_{intra} = 0.98$)
7. The average expert intra-agreement was 80.23% but of these, expert I stood out as being the most consistent (89.04%) and experts Q and J were the least consistent (74.16% and 73.18% respectively).
8. The intra-agreement of the plausible random numbers was not significantly different from that expected by chance (50.2%). This confirms the derivation of the expected chance agreement.

Chapter 6 summary.

Chapter 6 described the detailed mathematical and statistical techniques used to analyse the scores recorded by the experts and the system in the validation study (chapter 5, section 5.4). A novel method was designed to calculate the agreement between 2 time related, discrete sequences such that when they were identical they scored a maximum agreement of 100% and when there was no similarity, they scored 0. A partial agreement for different scores occurring at the same time was given as well as a partial agreement for similar scores with close temporal proximity. The application of this technique is not restricted to the analysis of the scores obtained from the validation study; the method has inherent flexibility which would allow it to be applied to any problem where it was desired to measure the similarity between discrete time related sequences.

The statistical distribution of the agreement measure was found to be normal with mean, $\mu_o = 52.7\%$ and variance, $\sigma_o^2 = 0.224$. This distribution was transformed into kappa statistics with $\mu_k = 0$ and $\sigma_k^2 = 0.998$ to test the statistical significance of a calculated agreement in excess of that expected by chance.

The Kruskal-Wallis one way analysis of variance by ranks was used to analyse the levels of agreement obtained by each review for the management of a given case. This test identified reviewers who obtained significantly lower agreements than the majority for their management of labour.

A synopsis of the results obtained by these tests can be found in the summary for chapter 5.

Chapter 6 references.

- Clark L, Mugford M and Paterson C. (1991) How does the mode of delivery affect the cost of maternity care ? *Brit. J. Obstet. Gynaecol.* **98**: 519-523.
- Cohen J. (1960) A coefficient of agreement for Nominal Scales. *Educ. Psychol. Meas.* **20**: 37-47.
- Cohen J. (1968) Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.* **70**: 213-220.
- Donker D K. (1991) Interobserver variation in the assessment of fetal heart rate recordings. *VU University Press, VU Boekhandel/Uitgeverij bv, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.*
- Fleiss J L. (1981) Statistical methods for rates and proportions. *John Wiley & Sons New York, USA.* 212-236.
- Hettmansperger T P. (1984) Statistical inference based on ranks. *J. Wiley, New York USA.*
- Kruskal W H. (1952) A nonparametric test for the several sample problem. *Annals math. stat.* **23**: 525-540.
- Kruskal W H and Wallis W A. (1952) Use of ranks in one criterion variance analysis. *J. Amer. Stat. Assoc.* **47**: 583-621.
- Landis J R and Koch G G. (1977) The measurement of observer agreement for categorical data. *Biometrics.* **33**: 159-174.
- Lehmann E L. (1975) Nonparametric statistical methods based on ranks. *Holden-Day, San Francisco, USA.*

Chapter 7

Discussion and Conclusions.

7.1	Review	218
7.2	Discussion of the system's performance	219
7.3	Discussion of the experts' performance	223
7.4	Some personal observations	228
7.5	Future development of the system	230
7.6	Looking further into the future	235
7.7	Final conclusions	238

7.1 Review.

This thesis began with an investigation of the factors which influence the quality of the fetal electrocardiogram (ECG), the primary signal from which heart rate is derived. This study found that the ECG was most seriously affected by low frequency baseline shifts. For heart rate calculations, which requires the accurate detection of the R-wave, these effects can be reduced with suitable bandpass filtering. The R-wave was found to have a bandwidth of 4 - 50 Hz. A bandwidth of 4 - 45 Hz reduced the R-wave amplitude by 3% but as this moves the upper frequency away from the mains frequency (50 Hz), it may be more desirable. Baseline shifts can become so severe that they cause saturation during pre-amplification prior to filtering which prevents signal analysis. It was found that the choice of fetal scalp electrode was important in reducing this possibility. The single spiral type was found to obtain signals with lower baseline noise and signal saturation more reliably than any other types considered (Copeland reusable, Copeland disposable and Hewlett-Packard double spiral). This finding has special importance for research work involving the analysis of ECG variables other than heart rate. The suggested bandwidth for the continuous recording of the ECG is 0.05 - 100 Hz. This lower limit means that baseline shifts cannot be easily removed from the signal with high pass filtering. However, it was argued that when analysis is concerned only with ECG complexes extracted from the continuous waveform, this lower frequency may be excessive. This may be especially true when successive ECG complexes are averaged.

The primary investigation of this thesis described the development, evaluation and validation of an intelligent system for labour management. The cardiotocogram, (CTG) provides a continuous visual recording of the fetal heart rate and uterine contractions from which an assessment of fetal condition during labour can be inferred. Considerable expertise is required to interpret the complex changes seen on this recording to accurately distinguish the fetus coping appropriately with the stress of labour from the compromised fetus. It is now accepted that difficulties in CTG interpretation can lead to unnecessary operative intervention and importantly, a failure to intervene when necessary. This is one of the major factors in infants who sustain potentially preventable neurological damage or mortality resulting from asphyxia during birth. Over the past 10 years, computer systems have been developed to attempt to provide quantitative heart rate analysis to overcome the inconsistencies of visual interpretation but have had little success. This is perhaps because the correct assessment of fetal condition depends not only on heart rate changes but also on considerable physiological knowledge, the specific patient history (mother and fetus) and expert clinical opinion. An intelligent system which embodies this expertise could be more successful.

The intelligent system conceived was developed closely with experts in Plymouth and during 2 limited internal evaluations, it was found that the knowledge required for labour management could be formalised and implemented and that the system could manage labour with a performance comparable with internal experts. The possible limitations of a single centre developing this type of system in isolation was recognised. The validation of the system compared the management of the system with 17 experts from 16 leading centres in fetal monitoring within the UK. Fifty cases were selected from a database of 2400 high risk labours. Each expert reviewed the cases twice, independently and blind to perinatal outcome at least 1 month apart. No expert had a knowledge of the other participating experts' identity. The objectives for this study were to investigate whether the;

1. System could manage labour with a performance comparable with the experts.
2. The experts could agree on the management of labour.
3. The experts could be consistent in the management of labour.

The evidence obtained from this study was both positive and conclusive. The system was found to be indistinguishable from the experts, except it was more consistent. The experts obtained good agreement with each other and reached a high level of consistency. This work demonstrates the potential for an intelligent system for labour management.

7.2 Discussion of the system's performance.

It was found that,

1. The system agreed with the experts well and significantly in excess of that expected by chance (67%, kappa = 0.31, $\alpha = 0.01$, $p < 0.0000003$).
2. In 45 cases, the level of agreement the system achieved with the experts was not significantly lower than the experts achieved with each other.
3. The actions recommended by the system were highly consistent (99%, kappa = 0.98, $\alpha = 0.01$) when used by two operators independently of which one was an engineer with no labour ward experience.
4. The system recommended delivery by CS in 11 cases. On average, more than 31 of the 34 expert reviews of these cases also recommended CS delivery. The majority of these did so within ± 1 segment (15 minutes) of the system's recommendation and two thirds did so within ± 2 segments (30 minutes).
5. The system identified as many of the cases which had birth asphyxia at delivery as the majority of experts and 1 more than was acted upon clinically.
6. The system was within the range of the experts interventions in cases in the poor outcome group (pH < 7.05).

7. The system recommended no unnecessary intervention in the cases with good outcome (spontaneous vaginal delivery and $\text{pH} > 7.20$) which was better than all but two of the experts.

However, although within the range of experts for all measures considered and therefore indistinguishable from the experts, the system did recommend fewer FBS and fewer CS than most. In addition, the system was different from the majority of experts in the management of four cases of which two would have been delivered with acidosis ($\text{pH} < 7.05$) and two would have had a significant metabolic acidosis ($\text{pH} < 7.05$ and $\text{BDecf} \geq 12$).

The examination of the system's management of these cases which obtained a poor outcome (chapter 5, section 5.5.8) found that unlike the majority of experts, the system did not react as quickly to signs of fetal compromise during the second stage of labour and did not interpret a rapid fall in baseline heart rate from slight tachycardia (160-180 bpm) to slight bradycardia (90-110 bpm) as the severely abnormal event that it was.

If acidosis is the condition to be avoided at birth (because there is some doubt that on its own it is) (Johnson et al, 1990) then it would be necessary to modify two rules in the system for it to achieve the management of the majority of experts in these four cases. The knowledge tree regarding baseline changes less than 90 bpm (chapter 4, section 4.3, figure 4.3), would need to be modified to;

- (i) **If** (Baseline < 90) **or** (Baseline > 180 and drops < 160) **or**
(Baseline > 160 and drops < 110) **Then**

activate this part of the tree.

- (ii) **If** (i) and (in second stage of labour) **Then**

recommend intervention.

It is of interest that these four cases with acidosis did not receive intervention in clinical practice. As these cases could be considered to have had a poor outcome, the question is, was it desirable that the system should have recommended intervention? Unfortunately, the answer is not clear. On the one hand, none of these babies had signs of neurological damage which may support the argument that intervention was unnecessary. On the other hand, the majority of babies who do sustain neurological injury do have acidosis. This implies that if acidosis at birth could be prevented then cases with neurological damage would be reduced. However, there are two mechanisms which cause two forms of acidosis, termed respiratory acidosis and metabolic acidosis. Respiratory acidosis is caused from an accumulation of CO_2 resulting from a reduction of blood flow across the placenta which is inhibited by contractions. An accumulation of CO_2 is normal and stimulates the baby to take its first

breath when born. But when the accumulation of CO₂ is severe, it can cause the baby to be born depressed. However, this situation can be quickly corrected and is not in itself an indication that the baby has become hypoxic (O₂ deprivation in organs) which is a requirement for asphyxia and neurological damage. Metabolic acidosis is caused by the production of lactic acid, which is produced when the fetus switches to anaerobic metabolism as a result of hypoxia. This process involves the breakdown of glycogen reserves stored principally in the liver and heart. Anaerobic metabolism is a natural defence mechanism but one that the fetus can only sustain for a limited period. Once the glycogen reserves are exhausted, then if hypoxia continues, tissue damage will occur and ultimately the fetus will be lost. So, should a significant metabolic acidosis be the condition at birth to be avoided? Anaerobic metabolism allows the fetus to survive the ordeal of a difficult labour after which the fetus will recover. Nevertheless, it seems dangerous to allow the fetus to approach the situation where its reserves become threatened and is put at considerable risk. In terms of the management of these cases, there is no way of knowing how much in reserve the fetus has which determines the maximum time the obstetrician has before an operative delivery is imperative. The problem is further complicated because even when the decision for operative delivery has been taken, it takes time to obtain. This argument suggests that a significant metabolic acidosis at birth should be avoided which is indicated when the base deficit in the extracellular fluid reaches 12 mmol/l (Siggaard Andersen, 1974). However, some care is required when assigning thresholds because there are always exceptions. The growth retarded fetus for instance may become compromised but have small reserves of glycogen and therefore may be incapable of sustaining anaerobic metabolism for any length of time. Consequently lactic acid may not be produced in sufficient quantities to indicate a significant metabolic acidosis by this definition.

Where does this leave the system? Of the four cases with acidosis, two cases had significant metabolic acidosis (cases 30 and 41). The system would have recommended intervention in both cases only if both of the proposed rule changes had been implemented. However, the inclusion of these rule changes would mean that the system would also have recommended operative delivery in the two cases with respiratory acidosis where intervention could be regarded as unnecessary. It is important that the system should be regarded as safe and so on balance it is likely to be considered that two cases operatively delivered with respiratory acidosis is an acceptable price to ensure that two cases with a significant metabolic acidosis are delivered in good time. The system should be perceived as slightly but not overly defensive and should err on the side of caution otherwise it will not instil confidence in clinical staff which is perhaps the most important factor in determining the eventual success of the system. These rules would be simple to implement and would not alter the management recommended by the system in any other case. The merits of these changes will be discussed with physiologists and obstetricians before they are implemented.

It was also found that the system did not recommend a CS in several cases where the majority of experts did, but where there was little evidence to suggest fetal compromise. Several of these cases were also delivered by CS clinically. It could be argued that as these cases did not result in a poor outcome that the experts' recommendations were inappropriate and that the system's recommendations are preferred. However, this could be a mistaken view. Consider the analogy that for the fetus, labour is like a steeple chase. The fetus enters the race as labour commences but the length of the race is unknown, as are the number of hurdles that are required to be jumped, represented by the contractions. These variables are determined in most cases by the mother and how quickly she progresses through labour. In some cases she can stop making progress in which case an end may not be possible without assistance. The obstetrician can influence the end of the race with the administration of drugs to speed up labour or by intervening operatively. The effect of these drugs is to increase the number of hurdles, make them higher and shorten the space between them. At certain points during the race, the obstetrician will make an assessment as to whether the fetus is likely to get to the end of the race in a reasonable condition. Many factors will weigh in the obstetrician's mind; where in the race we are, how quickly progress has been made to date, how much of the race remains, the present condition of the fetus and the estimated condition of the fetus if the race were allowed to run its course. The true expert will assess the situation with great skill. This analogy suggests that the obstetrician need not wait for the fetus to become compromised before an operative intervention is recommended and justified. This approach has the advantage that it gives time to plan the operative delivery without the need to resort to emergency procedures. The disadvantage is that it could be considered a gamble because unless one is particularly skilful, some women will receive operative intervention unnecessarily. To illustrate this, consider expert H; in case 15, expert H alone consistently recommended operative intervention, but the evidence she obtained from FBS results, which were also obtained by many of the other experts at the same time, did not suggest the fetus was as yet compromised. In clinical practice this case eventually delivered spontaneously but was birth asphyxiated. In case 22 however, expert H consistently recommended delivery by CS, again when the FBS results she obtained did not suggest fetal compromise. Here the gamble did not pay off because the baby actually went to be born spontaneously and had a good outcome.

The system has not so far been developed with this approach to labour management. Instead it waits for evidence of fetal compromise before recommending action. But it has been discussed that fetal decompensation can occur very rapidly and so in a way, the system is perhaps gambling too. It is assuming that if compromise does occur that clinical staff are on hand to expedite delivery quickly. It may be desirable for the system to be modified in a limited way to adopt an approach similar to the experts. However, the role for the system should principally be to identify fetal compromise but it could also warn staff when it considered it unlikely that a particular labour would achieve a good outcome at the current

rate of progress and stage in labour. The scenario for such a recommendation would be for example, a woman presenting early in labour considered to be high risk with an abnormal, but not severely abnormal CTG. In this case, it may be considered that there was little chance of obtaining a normal delivery with a normal outcome, and therefore little point in waiting for the CTG to deteriorate to severely abnormal before recommending delivery.

7.3 An assessment of the experts' performance.

This study also investigated whether nominated experts could agree and be consistent in the management of labour. It was discussed in chapter 1 that several studies have been undertaken to investigate agreement and consistency in CTG interpretation but few have been concerned with the intrapartum CTG and fewer still have incorporated specific case information. In addition, no previously reported study has obtained the complete relevant case information, devised a method for presenting the cases similarly to clinical practice and provided an option to obtain a fetal blood sample. For these reasons, this is the first study which seeks to investigate whether experts can agree and be consistent in the management of labour. This study found that;

1. All experts obtained an agreement with each other significantly better than could have been expected by chance. The average inter-agreement was 70%.
2. All experts were consistent significantly in excess of that which could have been expected by chance. The average intra-agreement was 80%.
3. There was significant ($\alpha = 0.1$) inter-expert disagreement in just 1 case (case 2).
4. In the cases recommended for CS by 14 experts, at least 24 of the other 32 expert reviews also recommended delivery by CS.
5. In the 31 cases where at least one expert recommended a CS, the majority of other experts who also recommended delivery by CS did so within ± 1 segment (15 minutes) and two thirds did so within ± 2 segments.
6. The majority of experts recommended operative intervention in 10 out of the 12 cases with acidosis ($\text{pH} < 7.05$) at delivery.
7. The majority of experts did not recommend operative intervention in cases which obtained a normal delivery with good outcome, except in 1 case in the second review where 9 out of 17 experts recommended intervention in the second stage.

However, the majority of experts did not recommend intervention in one case which was born birth asphyxiated. In addition there was disagreement and inconsistency in the cases recommended for FBS.

The majority of experts obtained results consistent with this summary, however, four experts obtained results that were conspicuous from the majority and were considered in greater detail.

Expert Q.

Expert Q was considerably different from the other experts in almost all respects. He scored 5 times as many segments with a protocol score of '4' than the average and 4 times fewer segments as '1'. This indicated that he was so concerned for the fetus that he was considering operative intervention 5 times as often as the other experts and he was only satisfied that there was no cause for concern for a quarter of the time compared to the average. If this expert had actually been managing these cases he would have found little time to relax and would have spent much time getting ready to whizz the bed out of the labour ward and into theatre. The most significant factor in the distribution of expert Q's scoring was that he never requested a fetal blood sample, which was his choice. The differences in his scorings perhaps would also explain why for 28 cases and 36 of his reviews, expert Q obtained a lower agreement for his management than the majority of the other experts. It could also explain why his inter-agreement with the other experts was the lowest and barely above that which could have been expected by chance (review 1, 59.5%, kappa = 0.14; review 2, 56.8%, kappa = 0.09). These points, by themselves do not indicate that expert Q managed these cases inappropriately. On the contrary, if he was able to attain a similar performance as the other experts in other respects, then there would be considerable evidence to suggest that he was able to extract all the information he required to manage labour from the CTG alone, without needing to intervene to obtain a FBS. This would demonstrate that he was different from the others by virtue of the fact that he was the most expert. But in fact, expert Q was the second most inconsistent expert in his scoring and was the most inconsistent in the cases he recommended for CS; in 9 cases, he either recommended a CS in his first review and not his second, or in his second review and not his first. He performed as well as the majority of experts in recommending intervention in cases with birth asphyxia, but he was more inconsistent and recommended lower intervention than most in cases with acidosis. He also recommended 2 unnecessary CSs and a second stage intervention in cases which clinically went on to have a normal delivery with good outcome.

To summarise expert Q's performance; he did not agree well with the other experts, was inconsistent, recommended CS intervention in cases with normal outcome but performed as well as the majority in cases with birth asphyxia. At first glance, one seems bound to conclude that expert Q is not an expert. The strongest evidence for this would seem to be his inconsistencies and interventions in cases with good outcomes.

Whilst it is possible that expert Q is not an expert, this does not seem likely because he is one of the most experienced of the experts; he is actively involved in teaching CTG interpretation, he has written books on the subject and gives frequent lectures. He is actively involved on the labour ward on a daily basis almost to the irritation of his juniors (personal admission). He is also considered eminent and is frequently asked for an expert opinion in cases which come to litigation.

There may be an alternative explanation for expert Q's performance which lies in the fact that he did not recommend a single FBS. The question is, why did he refuse to request a FBS when he knew it would obtain additional information to better identify the cases to deliver operatively? The answer is perhaps that he believed his powers of CTG interpretation were sufficiently good to remove the need.

The summary of expert Q's performance sounds remarkably similar to the established view regarding the current status of fetal monitoring discussed in chapter 1 which have led many to question the value of the CTG. However, in chapter 1, an alternative model was proposed to illustrate that although the CTG may be imprecise when taken in isolation, if used by an expert with detailed case information and an option to obtain further information from a FBS, it would tend to lead experts to agree and be consistent in their management by allowing the compromised fetus to be more accurately identified. This model predicted the findings of this study and significantly, it also predicts the results of expert Q. His experience and expertise in CTG interpretation were sufficient to identify the birth asphyxiated cases as well as most of the other experts. But his expertise in CTG interpretation was not a sufficient substitute for the additional information that could be obtained from fetal blood sampling. He was able to consistently identify the very compromised fetus but only at the expense of inconsistent and unnecessary operative interventions. By not obtaining additional information from a FBS his accuracy was reduced.

Importantly, if this model is correct then it predicts that a conventional computer solution to identify the compromised fetus during labour based on the CTG alone, will not be accurate. Expert Q performed less well than the other experts when considering the CTG together with patient information. A conventional computing approach which considered only the information obtained from the CTG, is likely to achieve a performance somewhat inferior to expert Q and would fall short of the performance demonstrated by the other experts and the system. This probably explains why previous attempts to computerise the CTG interpretation have not been successful.

Expert H.

Many of the results obtained for expert H were similar to expert Q. She obtained a relatively low agreement with the majority of other experts (review 1, 64.2%, kappa = 0.24; review 2, 69.8%, kappa = 0.36) and she recommended delivery by CS in two cases and second stage interventions in a further two cases which obtained a normal delivery with good outcome. Expert H also recommended the second highest number of CS deliveries. However, Expert H consistently recommended operative delivery in all cases which were delivered birth asphyxiated or with significant metabolic acidosis and consistently recommended operative intervention in all cases with respiratory acidosis except one which was identified in one review but not in the other. She was the most consistent expert in the cases she recommended for CS with only 1 out of 25 recommendations inconsistent.

As expert H was highly consistent, she demonstrates her expertise. However, because of her low inter-agreement, she is clearly working with different objectives than the other experts. Expert H can be regarded as a high interventionist. It seems she is working to a policy which regards a high CS rate as preferable to acidosis at birth. Many obstetricians would try to emulate this approach. Expert A for example, recommended a similar number of cases for CS delivery (21) but was inconsistent in 6 cases and only recommended operative intervention in 6 out of 12 cases with acidosis in his first review and 7 out of 12 in his second. It would be difficult indeed for many to implement the policy of high intervention to prevent acidosis with quite the accuracy of expert H. It is because of this effectiveness that many obstetricians would describe expert H as the most expert of the experts. She would have consistently avoided all but 1 of the cases with respiratory acidosis and all cases born with significant metabolic acidosis or birth asphyxia, which included case 15, for which only one other expert consistently recommended intervention.

Expert M

In contrast, expert M is of the low interventionist school. He recommended the fewest cases for CS and requested the fewest fetal blood samples (apart from expert Q!). Expert M consistently recommend no intervention in cases which obtained a normal delivery with good outcome. However, expert M would have allowed more cases to be born with acidosis than the majority but would not have allowed more cases to be born birth asphyxiated than the majority. For expert M, it seems the emphasis is on avoiding unnecessary intervention at the expense of some acidosis. It could be argued that expert M is the most expert of the experts because he recommends lowest intervention and he prevents the same number of birth asphyxiated cases as most of the other experts. However, some would argue with the same reasoning as that previously applied to the system, that

expert M is risking more than most by accepting a higher level of acidosis in preference to preventing unnecessary intervention.

Expert I.

Expert I is the experts' expert. She obtained the highest agreement for her management (review 1, 73.9%, kappa = 0.45; review 2, 74.7%, kappa = 0.46) and she was highly consistent (89.0%, kappa = 0.77). In addition, she received the second highest agreement for the cases she recommended for CS. The precision in her timing of CS recommendations between her 2 reviews was also remarkable because the average difference in her timings was less than 1 segment (15 minutes). She intervened in cases with poor outcome similarly to the majority of experts and only recommended 1 second stage intervention in a case with a normal delivery and good outcome. It is because she was the most consistent and is agreed with most often, that it could be argued that expert I was the most expert of the experts.

Expert I seems to represent the views of the majority of experts and it is of significance that it was this expert for which the system obtained its highest agreement (72.6%, kappa = 0.42).

The plausible random numbers.

The two sets of plausible random numbers were generated to investigate the results one would expect to obtain if the management of labour was not patient specific. The random scores were generated to have the same average distribution as the experts and obeyed the rules of the protocol. If the management of labour was not specific then it would be expected that random numbers would obtain results similar to the experts. If however, the management of labour was specific to the patient, then it would be expected that the random numbers would agree with each other no better than expected by chance and considerably worse than the experts agreed with each other. It was found that the random numbers;

1. Disagreed significantly with the experts (47.1%, kappa = -0.12, $\alpha = 0.01$).
2. Did not agree with each other significantly in excess of that expected by chance (50.2%, kappa = -0.05, $\alpha = 0.01$)
3. Were inconsistent in all but 1 of the 19 cases recommended for CS
4. Obtained a lower agreement than the experts obtained with each other in the cases recommended for CS.

These are the results one would expect from the experts if they were inconsistent and could not agree in the management of labour. Clearly these results are somewhat different than those obtained from the experts. The fact that the random numbers have not agreed significantly better or worse than expected by chance, supports the derived measure for the expected chance agreement described in chapter 6.

7.4 Some personal observations.

The evidence obtained from cases which result in litigation for injury or death sustained from birth asphyxia, suggests that for what ever reason, that the standards of fetal monitoring which can be reached, are not always obtained. The consequences can be tragic and made more so because they are sometimes preventable. This was the stimulus for one of the experts to suggest to me during the validation study that, "would it not be better that all high risk pregnancies should be delivered by caesarean section"? The suggestion was made to provoke a discussion and I willingly obliged. My first reaction was that, on the face of it, this appeared to be a rather shocking and crass suggestion, but financially the argument has credibility. The costs of one prevented litigation settlement would pay for approximately 1000 CSs. Indeed, this argument has manifested itself, albeit indirectly, in the USA where litigation is extensive and is feared by clinicians, with the consequence that as many as one third of all deliveries in some units are by CS (Banta and Thaker, 1979). However, despite the high CS rate, litigation is not falling (Symonds, 1991) and babies continue to be born damaged from birth asphyxia (Nelson, 1991). But the reason this argument seems so unreasonable is that it resigns itself to the conclusion that sub-standard monitoring is inevitable and cannot be improved in any way other than to avoid labour altogether.

It seems that fetal blood sampling is used to arbitrate in cases where the information obtained from the CTG is not clear. Some experts recommended fetal blood sampling considerably more than others which would seem to indicate that they were either more defensive or were not as confident in their interpretation of the CTG as those with low FBS rates. But could not a strategy be conceived whereby the need to record the CTG was removed and fetal blood sampling carried out periodically throughout high risk labour? Not for at least one of the experts involved in the study who described fetal blood sampling as a, "pain in the neck for midwives, a pain in the back for doctors, a pain in the head for the baby and a pain somewhere else for the mother." Few would deny that fetal blood sampling is undesirable and its use should be minimised which again would be addressed in part, if labour ward staff were highly skilled in CTG interpretation.

7.4.1 Improving the standards of fetal monitoring.

The results obtained from the validation study would seem to suggest that the standards of fetal monitoring can be improved. If this is the case then how could this be done?

1. More senior staff could become more involved on the labour ward; their role would perhaps be to make themselves aware of the cases currently in labour and to oversee that the appropriate standard of care was being maintained. They could do this discretely if central monitoring techniques were employed, which are available and allow cases to be reviewed without further interfering with the patient. This may help because it could reduce the responsibility given to junior staff who are often those closest to the patient and are required to form a judgement as to whether more senior staff should be called. When visiting an expert's home during the validation study, I was introduced to his method for keeping tabs on the labour ward; his fax machine. He demonstrated to me that he was able to telephone the clinician on the labour ward and ask for a patient's CTG to be faxed to him which allowed him to discuss the case. This technique has also been suggested by one of the Consultants in Plymouth but the idea was not taken up.

In at least one of the hospitals I visited, the routine care of all patients in labour was transferred to one of two Consultants with recognised expertise. Either one of the Consultants was present on the labour ward and could deal with any complications as they arose. The involvement of more senior staff on the labour ward is likely to help improve standards but it seems unlikely that this expertise will be available for all monitored births day and night.

2. Better training of medical staff; currently, clinical staff are not formally trained in CTG interpretation. Junior midwives and junior doctors are expected to learn from those more experienced who in turn have relied on the informal guidance of others. It would seem, would it not, that in some instances this process may result in the blind leading the blind, leading the blind. With this approach, it is difficult to assess the quality of the training staff have received which may range from anywhere between excellent to non-existent. A more consistent approach to ensure that all receive an adequate standard of training would be to formalise training in CTG interpretation prior to their involvement on the labour ward. This is likely to improve standards, but if CTG interpretation is difficult then it may not be possible for all clinical staff to fully master the techniques required.
3. Intelligent systems to advise on labour management; the possibility has been demonstrated in this study that current CTG recorders could one day, be transformed

into true fetal monitors which could assist clinical staff in their decision making. This is supported by the validation study where it was found that an engineer guided by the system obtained a performance comparable with the experts. The role of such a system could be as a safety net to address the limitations of the proposals discussed above and thereby provide a high standard of monitoring both day and night.

7.5 Future development of the system.

7.5.1 CTG Feature extraction methods.

The devised feature extraction methods enabled the system to obtain a performance in labour management comparable with experts. This was because the methods were developed closely with experts and were designed to be flexible to allow the experts to shape the classification criteria. This was a rather different approach to the problem than the conventional computing approaches where the classification criteria were made without reference to the experienced clinician but using thresholds derived empirically. There are however, some areas of feature extraction which could be improved.

Further development of the baseline algorithm is recommended because its correct interpretation is key to the interpretation the other heart rate features. The rare problem of the unstable baseline discussed in chapter 3 (section 3.7) cannot be reliably solved by the baseline algorithm in its current form. There are several ways in which this problem could be tackled. The first step would be to make the algorithm aware of when the unstable baseline situation exists. This would be characterised when the distribution of heart rate samples become bi-modal, indicating two baselines were possible. The algorithm must then decide which is the correct modal heart rate to expect, the upper or the lower value? This dilemma could be resolved from an examination of previous segments of recording when the baseline was stable. The modal value closest to this stable value could then be taken. But resolving the problem of an unstable baseline at the commencement of monitoring where previous analysis was unavailable, would also need to be considered.

An alternative approach may come from the fact that during unstable periods, technically there is no baseline heart rate and as such it is an abnormal event. It could be that the most abnormal classification that could be made from the trace would be the most appropriate to use. Alternatively, the algorithm could be developed to recognise unstable baselines and simply classify these periods as abnormal and refer the expert system to the knowledge tree which handles 'abnormal' traces.

whether the trend is gradually upwards or downwards. This may give additional information regarding how the fetus was responding.

Present methods obtain a measure of heart rate variability during stable periods of recording between accelerations and decelerations within a 15 minute segment. However, this requires that at least one minute of stability exists, which is a fair assumption for most cases (indeed no case has been reviewed where this has not occurred). However, important information is also contained in the variability within decelerations. If a curve fitting algorithm were applied to fit the decelerations, the heart rate samples could be transformed upon a zero baseline from which a measure could be obtained using the current method.

The system currently identifies the location of contractions for the classification of heart rate decelerations only. But the location of contractions are also important for monitoring the administration of oxytocins, the drugs used to stimulate the uterus and augment labour. A problem associated with these drugs is that they can be over administered which leads to contractions becoming too frequent. This is particularly stressful for the fetus which relies on the period between contractions for recovery. If a fetus got into trouble as a result of hyperstimulation, then the system in its current form may advise an operative delivery. In practice the clinician may feel that the fetus would be able to recover if the drug dosage were reduced or stopped. This in turn, may possibly allow the fetus to be delivered normally.

As the system already obtains the information required to monitor contractions, it would be a relatively simple matter to extend the knowledge base to monitor the frequency of contractions.

7.5.2 Monitoring the progress of labour.

The progress of labour is assessed from the rate at which the cervix dilates which is normally 1cm per hour. This information is plotted on a graph by the midwife to enable an assessment of the woman's progress to be made. Sometimes, a woman can fail to make adequate progress which may be corrected with the use of drugs as previously discussed.

The current cervical dilatation and progress of labour are also important for determining the appropriate management when there are concerns for the fetus. This was illustrated in the steeple chase analogy. In addition, the validation study found that the second stage of labour, especially when the mother was pushing, influenced the management of the experts. They were much more inclined to request intervention when signs of fetal compromise were

apparent. The system did not make a distinction between the first and second stages of labour and was consequently slower to react.

These findings have identified an important area for the system to be extended and indicate that the management of labour cannot be regarded as discrete tasks. It seems it is a complete package which relies on information obtained from a variety of sources including the CTG, fetal blood sampling, progress of labour and administration of drugs etc, all of which are inter-related and collectively influence management.

7.5.3 On-line development.

The system has been developed to function using previously recorded data and as such operates off-line. It considers cases retrospectively and handles processing demands sequentially which is not appropriate for operation on the labour ward. To operate on-line, the system will be required to handle processing demands in 'parallel'. For example, a clinician may wish to obtain information from the system which may take several minutes. During this time data will have been collected and it is important that it is analysed quickly.

This represents a significant development but the development time could be reduced if the system were implemented in a programming environment which already managed processes in 'parallel'. The most successful of these is Microsoft Windows, for which versions of the most popular programming languages, such as 'C' have been adapted to be used with. This approach would mean that the system's algorithms would require little further modification. The main area for development would be the user interface and explanation facility. The feasibility of this approach has been demonstrated in a separate project undertaken by our group in partnership with a commercial company. Here an expert system was developed to interface to a blood gas analyser to provide an interpretation of the measurements obtained and to store this information in a database. This system is currently on trial in the clinical environment and is proving to be very successful.

7.5.4 Commercial exploitation.

If the system can be demonstrated to reduce unnecessary intervention and reduce birth asphyxia on the labour ward then its potential impact world wide could be enormous. It could reduce the numbers of damaged babies and could save Health Departments millions in prevented compensation payments and also make savings by reducing unnecessary CSs which can cost up to £1560 per delivery (Clark, 1991). The system could also be of considerable value to underdeveloped countries where fetal monitoring expertise is

unavailable. However, the potential benefits do not minimise the obstacles which could prevent the system becoming viable.

The effectiveness must first be demonstrated in the clinical situation during a carefully controlled randomised trial. This would compare clinical practice with and without the system in an independent centre. This remains the acid test for the system. These trials are expensive and would be difficult to fund without a commercial partner.

The system will also need to be acceptable to the user, because ultimately, if the clinical staff do not like the system then it will not matter how good it is, it will fail. Furthermore, there is likely to be a natural reticence towards any system which purports to be 'intelligent'. This could be overcome if handled with care, as the role for the system should be regarded as an advisor to the clinician who would remain in control. This can be achieved with the current system design because the knowledge has been represented into rules. When an action is suggested, the clinician will be able to ask why? and the system would be able to explain. If the clinician found fault in the systems explanation then they would be free to follow their own inclination.

A UK company who have experience in fetal monitoring have expressed an interest in the system. They manufacture antepartum and intrapartum CTG recorders and have recently released a personal computer based central monitoring system which also archives clinical record information. This system runs on a network which means that our system could be easily integrated and would have access to all the information it would require. This network approach seems the most suitable implementation for the system even if this particular commercial venture was not forthcoming. A patent application has been made.

7.5.5 A suggested work plan for the immediate future.

A proposed work plan which will take the system from its current position to be ready for a clinically controlled randomised trial in an external centre is represented in figure 7.1.

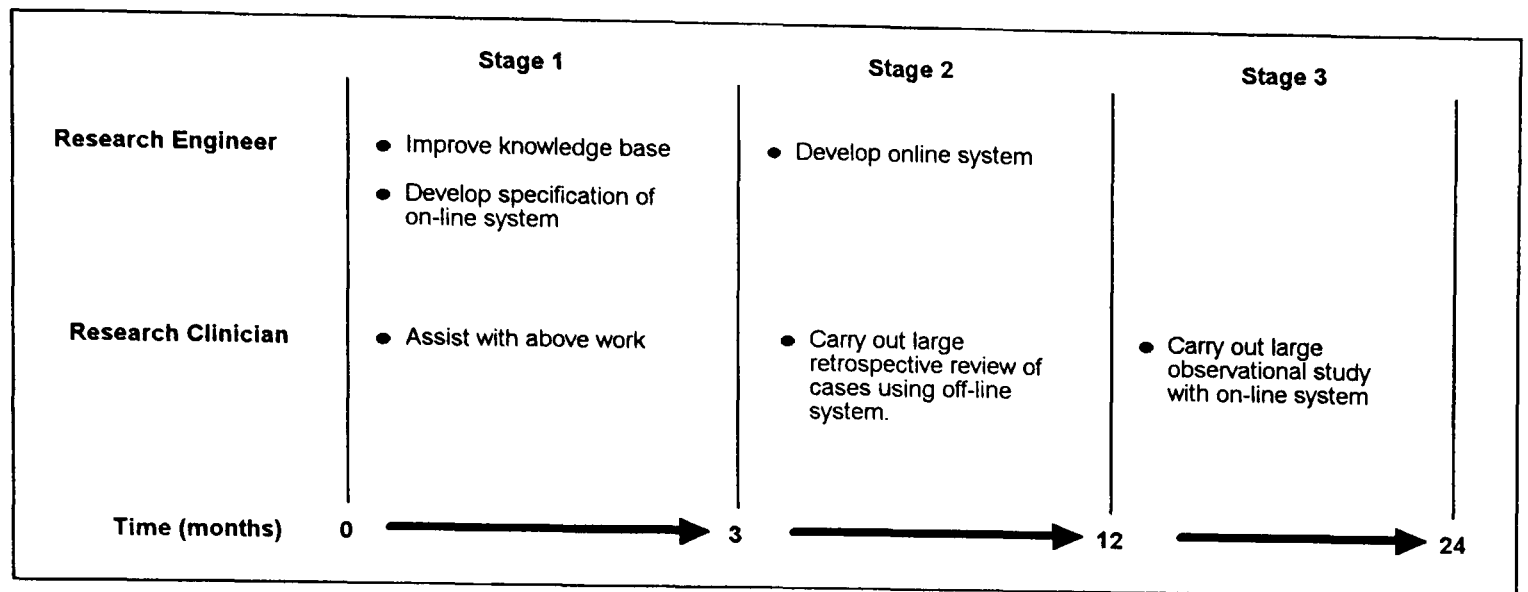


Figure 7.1. Proposed future work programme.

Stage 1. Extend the system in accordance with the findings of the validation study. The process of incorporating the changes in the embodied knowledge will require assistance from an experienced research clinician to reformulate the rules which make up the knowledge structure.

Next, the functional requirements for the on-line system will be obtained. This system will need to be reliable and user friendly to allow non-technical staff to interact with it efficiently.

Stage 2. The aim is to further develop the system to make it capable of working reliably *on-line*, in the labour ward. The user interface will be developed which is a requirement for stage 3.

In parallel, a research clinician should undertake a study to examine the performance of the system using many cases to fully assess the limitations of the system and identify improvements which the engineer can incorporate. A large number of cases are required to ensure that the system encounters sufficiently diverse examples for it to be fully tested. The cases collected which are considered difficult to interpret could be presented to external experts for review. In addition, difficult cases could also be obtained from other hospitals. These are not required to have been previously digitised, as a method, described in chapter 5 (section 5.4.5), has been devised to extract the data from paper traces.

Stage 3. This investigation would carefully examine the performance of the system on the labour ward in a trial lasting a year where approximately 900 cases would be monitored. The *on-line* system could function with its recommendations blind to clinicians. Each case could be subsequently analysed independently during the study by the research clinician and any difficult traces could be referred to external experts. On completion of the study the clinician could examine the fetal blood sampling rates and operative deliveries (actual and recommended) and compare these to perinatal outcome to see how the system was likely to have effected clinical decision making.

This information would then be used to assess whether the system was viable and likely to be acceptable to clinical staff. If the results from this study were promising then the system would be ready to undergo a clinical randomised trial.

7.6 Looking further into the future.

There are currently two areas of current research in Plymouth which may have implications for the future development of the system.

7.6.1. New methods for fetal monitoring.

Research has been underway for some time to identify a single variable which accurately indicates fetal condition that can be obtained reliably and with relative ease during labour. If such a variable were found it would have the potential to revolutionise fetal monitoring. But the inaccessibility of the fetus means that it is difficult to obtain physiological measurements with any degree of reliability. The existence of a single variable to indicate fetal condition seems unlikely to be found, which means the CTG is here to stay (Neilson, 1993). It has been discussed that the CTG, by itself is not accurate. The addition of fetal blood sampling improves accuracy but has undesirable properties. Its principal limitation is that it is intermittent. The pH can fall rapidly which could be missed, especially in the second stage of labour.

The oxygen saturation of the blood can be measured from changes in electromagnetic radiation as it passes through body tissue. This technique is routinely used to measure the oxygen saturation in patients during surgery and its application to the fetus is being investigated (Johnson, 1991). However, the location and inaccessibility of the fetus presents this method with certain difficulties, principally, what do you shine the electromagnetic radiation through? An electrode has been developed which can be positioned along side the fetal head. A light source is transmitted into the fetal head of which a proportion of

radiation is reflected back to a sensor located next to the source. The difficulties with this method involve ensuring the electrode remains correctly positioned and that the signal source does not shine directly into the sensor. At present these difficulties mean that the technique is too unreliable for routine clinical use and its effectiveness has yet to be demonstrated in a randomised trial. If this method is to be used with the CTG, then it also has the disadvantage of requiring an additional electrode to be introduced via the vagina.

The variable which is currently showing greatest promise is the ST-waveform of the fetal ECG and the changes which occur in it during hypoxia, (Greene, 1987).

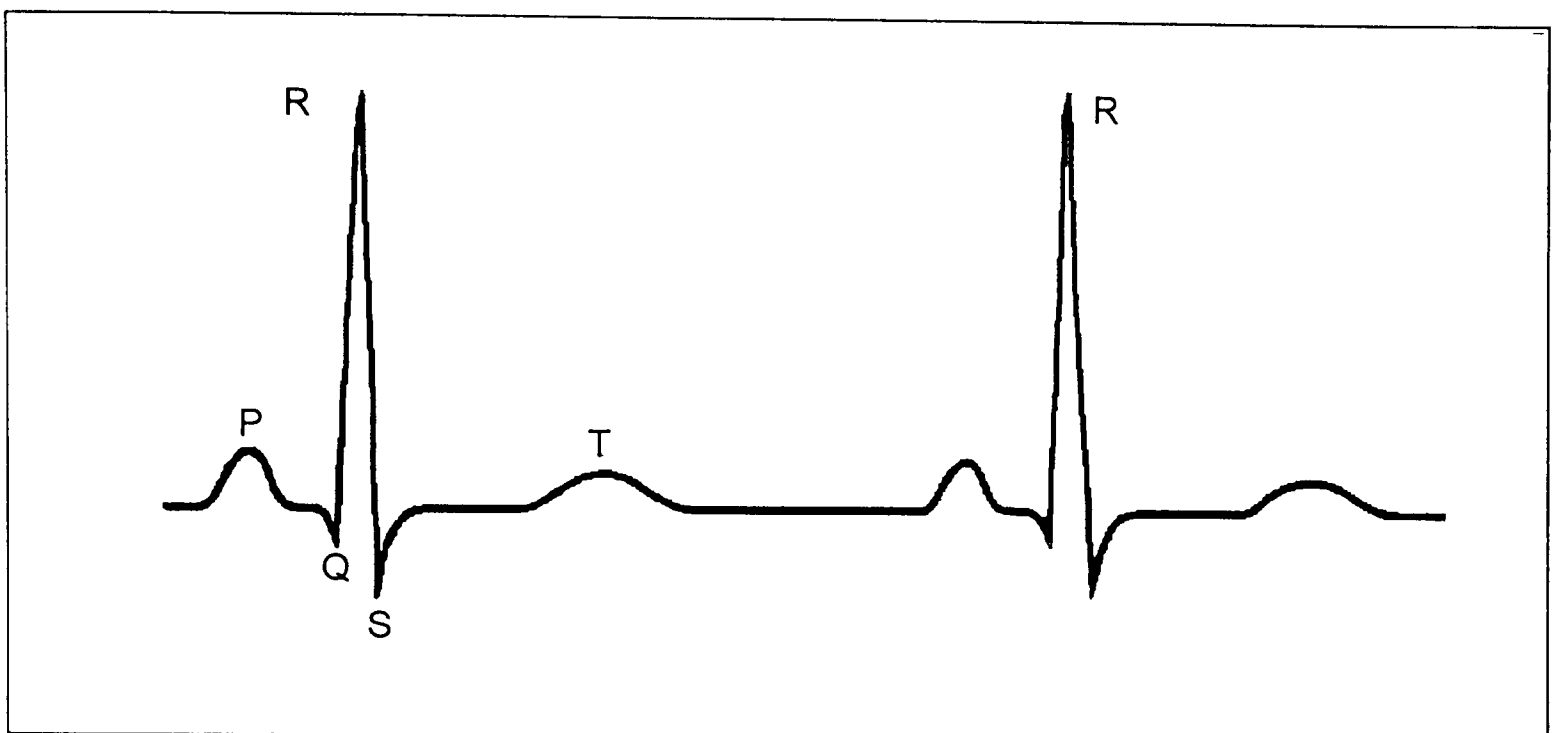


Figure 7.2: The ECG waveform.

The ECG waveform is shown in figure 7.2 and has been described in chapter 1. Like the adult, the fetal ECG is made up of the P, Q, R, S, and T waves. The ST waveform and T wave represent the active phase of the cardiac cycle and involves the repolarisation of the heart cells ready for the next stimulus. The adult ECG is used routinely for diagnosis purposes and in particular for stress tests during exercise. It is during these times of stress that changes occur in the ST waveform. It has been found that in the fetus too, this can provide additional clinically useful information.

It has been discussed that when the fetus becomes hypoxic it can switch to anaerobic metabolism by breaking down glycogen reserves. The by-product of this process is lactic acid but is also potassium ions. These can cause changes in the heart muscle cell membrane potentials which alter the shape of the ST waveform. Therefore changes in the ST waveform can provide continuous information on the metabolic functioning of the fetus.

The main advantage with the ST waveform is that it is available from the same signal source as that used to obtain heart rate. The processing of the ST waveform is computationally more demanding than heart rate but as was shown in chapter 2, if a single spiral electrode is used and precautions are taken in the design of the data collection system, then these difficulties can be overcome.

The clinical effectiveness of this technique has also been demonstrated in a randomised trial which compared conventional CTG monitoring, with CTG monitoring together with ST waveform analysis, (Westgate, 1992). This study was able to show that by including an analysis of the ST waveform, operative intervention was reduced by half without affecting outcome.

The advantage of the rule based approach adopted for the system is that the knowledge can be easily modified and updated to include information obtained from new techniques.

7.6.2 Reasoning with uncertainty.

A possible limitation with the current form of the systems knowledge is that it is 'crisp'; it relies on facts which can either be true, or false. Measures in the natural world are seldom in this form, they are usually shades of grey.

Research is underway in Plymouth, to investigate whether the performance of the system could be improved by incorporating inexact or approximate reasoning. The most promising method for representing this approach seems to be, fuzzy set theory or fuzzy logic as it is often called (Zadeh, 1975a, 1975b, 1975c). This theory considers membership functions to assess how much evidence there is that a set of conditions are associated with a certain grouping. For example, the system in its current form considers the heart rate to be a tachycardia when it is greater than 160 bpm. This means that it would consider 161 bpm as a tachycardia, but would consider 159 bpm as normal. Fuzzy theory takes a different view. It would regard a baseline heart rate of 175 bpm to almost certainly be a member of the fuzzy-set of tachycardias. It would regard a baseline of 161 bpm as having significant but less evidence that it was a member of this set and similarly some, but less evidence for a baseline of 159 bpm.

7.7 Final conclusions.

The findings of this study were both positive and conclusive and as such challenge the view which questions the value of the CTG. It was recognised that the CTG when considered in isolation may indeed be an inaccurate test of fetal condition, but this study has shown that when experts apply their physiological knowledge and clinical experience and have access to all the relevant information, they can distinguish the compromised fetus from the fetus coping appropriately with labour with a high degree of accuracy.

The epigram at the start of this thesis is the conclusion to which it has been brought because for fetal monitoring, "the world of reality *is* a world of limitations". A single variable is unlikely to be found which accurately measures the condition of the fetus and is simple to obtain and reliable. However, this study has shown that despite the current limitations, expertise is achievable and suggests that the problems associated with fetal monitoring are more related to the transfer of knowledge to clinicians rather than the techniques they employ. The challenge remains to formulate a method to effectively transfer expertise to the labour ward and thereby address the real and practical problems which face fetal monitoring today. This thesis demonstrates that intelligent systems may provide the vehicle to achieve this.

Chapter 7 references.

- Banta H D and Thacker S B. (1979) Assessing the costs and benefits of electronic fetal monitoring. *Obstet Gynecol Surv.* **34**: 627-642.
- Clark L, Mugford M and Paterson C. (1991) How does the mode of delivery affect the cost of maternity care? *Brit J Obstet Gynaecol* **98**: 519-523.
- Greene K R. (1987) The ECG waveform. In: Ballière's clinical obstetrics and gynaecology. **1**: 1, 131-155.
- Johnson J W C, Richards D S, Wagaman R A. (1990) The case for routine umbilical blood acid-base studies at delivery. *Am J Obstet Gynecol.* **162**: 621-625.
- Johnson N, Johnson V A, Fisher J, Jobbings B, Bannister J and Lilford, R., J. (1991) Fetal monitoring with pulse oximetry. *Br J Obstet Gynaecol*; **98**: 36-41.
- Neilson J P. (1993) Cardiotocography during labour; An unsatisfactory technique but nothing better yet. *BMJ* **306**: 347-348.
- Nelson K N and Leviton A. (1991) How much of neonatal encephalopathy is due to birth asphyxia? *Am J Dis Child.* **145**: 1325-1331.
- Siggaard Andersen, O. (1974) The acid base status of the blood. Munksgaard, Ed. **4**.
- Symonds E M. (1991) Litigation and birth related injuries. In: How to avoid medico-legal problems in obstetrics and gynaecology. Chamberlain G (ed) RCOG, Chameleon Press, London, UK.
- Westgate J, Harris M, Curnow J S H and Greene, K., R. (1992) Randomised trial of cardiotocography alone or with ST waveform analysis for intrapartum monitoring. *Lancet*; **340**: 194-198.
- Zadeh L A. (1975a) The concept of a linguistic variable and its application to approximate reasoning - I. *Information sciences*, **8**: 199-249.
- Zadeh L A. (1975b) The concept of a linguistic variable and its application to approximate reasoning - II. *Information sciences*, **8**: 301-357.
- Zadeh L A. (1975c) The concept of a linguistic variable and its application to approximate reasoning - III. *Information sciences*, **9**: 43-80.

Appendix A

C - language software for analysis of the fetal electrocardiogram.

/* ANALYSE */

/* Program to assess the quality of ECG records stored on optical disk */
/* collected using different scalp electrodes. */

/* Written by Robert D.F. Keith */

```
#include <stdio.h>
#include <stdlib.h>
#include <bios.h>
#include <graphics.h>

#include <math.h>
#define CLS printf("%c[2J",27)

/* ANSI escape sequences to set */
/* up the clear screen command & */
/* the text, background colours */

#define SETCOLOUR printf("%c[1m%c[44m%c[36m",27,27,27)
#define INPUTCOLOUR printf("%c[1m%c[44m%c[37m",27,27,27)
#define RESET printf("%c[0m",27) /****** reset the screen *****/

#define BLUE_SCREEN setbkcolor(9)
#define WHITE setcolor(15)
#define LT_BLUE setcolor(9)
#define CYAN setcolor(11)
#define RED setcolor(4)
#define YELLOW setcolor(14)
#define GREY setcolor(7)
#define SOLID_LINE setlinestyle(0, 0, 1)
#define DASHED_LINE setlinestyle(3, 0, 3)
#define THICK_LINE setlinestyle(0, 0, 2)

/* graphics screen definitions ****/
/* setcolor sets the draw colour **/
/* setbkcolor sets screen colour **/

/* NOTE middle 0 has no effect **/
/* for these linestyle options **/
/* definition for use with flags*/

#define TRUE 1
#define size 15120

FILE *fp;
FILE *rwave;
FILE *base;
FILE *mains;
FILE *power;
FILE *snr;
FILE *spike;
FILE *drop_sat;

FILE *totals;

int data[15120], *pa;
int hours, minutes, seconds, hrs, mins, secs, g_driver, g_mode;
int file_flag;

int saturation_total;
int num_20sec_blks, num_5sec_blks;
int total_num_20sec_blks = 0;
int total_num_5sec_blks = 0;
float saturation_so_far = 0;
int drop_out;
float drop_out_so_far = 0;
int number_spikes;
int spikes_so_far = 0;
int power10;
int power20;
int power30;
int power40;

float stage1_coeff[12], stage2_coeff[70], decimated_data[4000],
      mains_coeff[72], baseline_coeff[96], rwave_coeff[72], x2[100];

float temp_array[500], mains_power, bline_power, rwave_power;
float baseline_snr, mains_snr;
```

```

float time = 0;

char file_name[30], scale[50];

/***** GLOBAL VARIABLES *****/

/* INT */
/* hours, minutes seconds = time of data section within the data array *****/
/* hrs, mins, secs is the time associated with the display scales *****/
/* g_driver & g_mode are variables which, after using detectgraph(), *****/
/* correspond to the graphics driver and mode set on the computer *****/
/* file_flag is set when file pointer is at eof *****/

/* CHAR */
/* file_name is the string containing the name of the file *****/
/* scale is the string holding the x axis time scale and is variable *****/
/* size is the constant associated with how much data is loaded into data **/
/* array. This represents just over 30 secs (30.24) for display purposes **/

/* FLOAT */
/* stage1_coeffs & stage2_coeffs are arrays holding the coefficients for **/
/* the decimation process. Decimated_data is an array holding the results **/
/* of the decimation, ie the decimated data. mains_coeff, baseline_coeff, **/
/* and rwave_coeff are the arrays which hold the coefficients which extract*/
/* the mains_component, baseline and rwave signal respectively from the ****/
/* decimated data. x2 holds the results of filter operations - a scratch */
/* array if you like. Tmp_array holds the filtered data ie either the mains*/
/* extracted data, or the baseline data or the rwave filtered data prior to*/
/* to calculation of the signal to noise ratios. */

/*****

/***** MAIN PROGRAM *****/

main()
{
int option, flag, i, analysis_flag, fd, g, loops, s, jump;
char alpha, beta;
int v = 0;
float x_axis;
long old_time = 0;
long new_time = 0;

assign_coeffs();          /** read decimation coeffs array */
BEGINNING:                /** for use with goto statement */

SETCOLOUR;                /** set up the screen *****/
get_file_info();          /** prompt for file info *****/
                           /** seek the data start position **/
get_data();               /** load raw data in array "data" **/

if(file_flag == TRUE)     /** if an eof encountered *****/
{
    errors(1);            /** print error message *****/
    goto BEGINNING;       /** go back to BEGINNING *****/
}

RESET;                    /** reset the screen *****/
detectgraph(&g_driver, &g_mode);
                           /** finds out the graphics driver */
initgraph(&g_driver, &g_mode, "");
                           /** & sets up screen accordingly **/
g_mode = EGAHI;
display_data();           /** display data *****/
getch();                  /** wait for a key to be pressed */
closegraph();            /** close down the graphics **/
flag = 0;                 /** flag set on exit option *****/
analysis_flag = 0;
while(flag != TRUE)
{
    SETCOLOUR;            /** set cyan text *****/
    CLS;                  /** set blue screen *****/
    file_flag = 0;
    option = options();
    if(option == 'z')
        /** zoom data *****/
        {
            initgraph(&g_driver, &g_mode, "");
            display_data();
            zoom_data();
        }
}
}

```



```

        closegraph();
    }
    if(option == 's')
        /****** scroll data *****/
        {
            initgraph(&g_driver, &g_mode, "");
            g_mode = EGAHI;
            scroll_data();
            if(file_flag == TRUE)
                /****** loop until quit *****/
                {
                    closegraph();
                    fclose(fp);
                    errors(2);
                    hours = 0;
                    minutes = 0;
                    seconds = 0;
                    fp = fopen(file_name, "rb");
                    get_data();
                    continue;
                }
                /****** display next 30 secs *****/
            closegraph();
        }

    if(option == 't')
        /****** time jump *****/
        {
            fclose(fp);
            fp = fopen(file_name, "rb");
            seconds = 0;
            jump = (60 * hours) + minutes;
            for(s = 0; s < (2 * jump) + 1; s++)
            {
                get_data();

                if(file_flag == TRUE)
                {
                    errors(3);
                    hours = 0;
                    minutes = 0;
                    seconds = 0;
                    fclose(fp);
                    fp = fopen(file_name, "rb");
                    get_data();
                    closegraph();
                    goto BEGINNING;
                }
            }
            RESET;
            initgraph(&g_driver, &g_mode, "");
            display_data();
            getch();
            closegraph();
        }
    if(option == 'v')
        {
            closegraph();
            v = 0;
            while(TRUE)
            {
                for(i = (v*500); i < (500*(v+1)); i++)
                    printf("%d\n", data[i]);
                beta = getch();
                if(beta == 'q')
                {
                    break;
                }
                else
                {
                    v++;
                    continue;
                }
            }
        }
    if(option == 'c')
        /***** new data file *****/
        {
            fcloseall();
            goto BEGINNING;
        }
        /****** get new file info *****/
    if(option == 'a')

```

```

{
printf("how many sections to be analysed\t");
scanf("%d", &loops);
old_time = biostime(0, old_time);
CLS;
    initgraph(&g_driver, &g_mode, "");
file_flag = 0;
g = 0;
power10 = 0;
power20 = 0;
power30 = 0;
power40 = 0;
x_axis = 0;
open_files();
do
    {
cleardevice();
display_data();
saturation();
signal_drop_out();
rwave_power = 0;
baseline_snr = 0;
bline_power = 0;
mains_snr = 0;
mains_power = 0;
if(saturation_total == 0 && drop_out == 0)
    {
delete_line(10);
YELLOW;
outtextxy(150,10, "Decimation in progress");
decimate();
filter_coeffs();
delete_line(10);
YELLOW;
outtextxy(150,10, "Mains extraction in progress");
extract_mains();
delete_line(10);
YELLOW;
outtextxy(150,10, "Baseline extraction in progress");
extract_baseline();
delete_line(10);
YELLOW;
outtextxy(150,10, "R wave extraction in progress");
extract_rwave();
delete_line(10);
YELLOW;
outtextxy(150,10, "Power calculations in progress");
calculate_mains_power();
calculate_bline_power();
calculate_rwave_power();
calculate_sig_to_noise();
if(bline_power <= 10)
    {
power10++;
}
if(bline_power > 10 && bline_power <= 20)
    {
power20++;
}
if(bline_power > 20 && bline_power <= 30)
    {
power30++;
}
if(bline_power > 30)
    {
power40++;
}
}
detect_spikes();
results_to_file(x_axis);
set_up_scales();
g++;
x_axis = x_axis + 0.5;

if(g == loops)
    {
analysis_flag = TRUE;
}
}
}

```

```

        }
        get_data();
        if(file_flag == TRUE)
        {
            analysis_flag = TRUE;
        }
    }
    while(analysis_flag != TRUE);

    new_time = biostime(0, new_time);
    time = new_time - old_time;
    totals = fopen("\\totals.rdf", "w");
    totals_to_file();
    fcloseall();
    closegraph();
    RESET;
    CLS;
    exit(1);
}

if(option == 'e')
    {
        flag = TRUE;
        closegraph();
        RESET;
        CLS;
        fcloseall();
    }
}

/***** exit program *****/

/***** reset screen to B/W *****/

/***** closeall open streams *****/

/***** GET FILE INFO *****/
get_file_info()
{
    RESET;
    while (TRUE)
    {
        CLS;
        SETCOLOUR;
        CLS;
        printf("\n\tName of STAN data file (full DOS path)? \n");

        INPUTCOLOUR;
        printf("\n\t");
        scanf("%s", &file_name);

        if ((fp = fopen(file_name, "rb")) == NULL)
        {
            errors(4);
            continue;
        }
        else
        {
            break;
        }

        SETCOLOUR;
        CLS;
        hours = 0;
        minutes = 0;
        seconds = 0;
    }
    /***** GET DATA FROM STAN FILE *****/
    /** This function will get bytes from a STAN file, strip out the markers **/
    /** and store the data in an array length 30000 (30 seconds worth) *****/

    get_data()
    {
        int sample, offset;
        /* sample = current data value ****/

```

```

pa = &data[0];
offset = 0;
START:
sample = get_sample();
if(file_flag == TRUE)
    {
        return;
    }
if (sample != 255) goto START;
sample = get_sample();
if(file_flag == TRUE)
    {
        return;
    }
if (sample != 1) goto START;
do
    {
        sample = get_sample();
        if(file_flag == TRUE)
            {
                return;
            }
        if (sample == 255)
            {
                goto START;
            }
        else
            {
                *(pa+offset) = sample;
                offset++;
            }
    }
while (offset != size);
/*****
/***** GET SAMPLE *****/
/***** Will get the next 8 bit data sample from the input file *****/
get_sample()
{
    unsigned int d_num;
    int test;
    test = feof(fp);
    if(test != 0)
        {
            file_flag = TRUE;
            fcloseall();
            closegraph();
            return(0);
        }
    d_num = getc(fp);
    return (d_num);
}
/*****
/***** ERRORS *****/
/***** function to print out an error message *****/
errors(int num)
{
    SETCOLOUR;
    CLS;
    if(num == 1)
        {
            printf("\n\tError 1: End of file in %s. Please re-enter\n",file_name);
        }
    if(num == 2)
        {
            /* scrolling takes pointer past eof
*/

```

```

    {
        printf("\n\tError 2: You have scrolled to the end of the file\n");
    }

if(num == 3)
*****
    {
        printf("\n\tError 3: Incorrect time entry: The file is not that long\n");
    }

if(num == 4)
**/
    {
        printf("\n\tError 4: File %s does not exist\n",file_name);
    }
INPUTCOLOUR;

printf("\n\tPress any key to continue");
getch();
}
/*****
/***** DISPLAY DATA *****/
display_data()
{
int x_cord,y_cord;
char option;

set_up_screen();
moveto(50,300-data[0]);
draw(1,28);
display_scales(30, 0);
WHITE;
outtextxy(250, 310, "Time [hr:min:sec]");
outtextxy(26,30, "No. Bits");
}
/*****
/***** ZOOM DATA *****/
zoom_data()
{
int zoom_area;

delete_line(400);
select_option();
scanf("%d", &zoom_area);

cleardevice();
set_up_screen();
delete_line(310);
moveto(50,300-data[0+((zoom_area-1)*3000)]);
draw(zoom_area,4);

display_scales(6, zoom_area);
WHITE;
outtextxy(250, 310, "Time [hr:min:sec]");
outtextxy(26,30, "No. Bits");
getch();

}
/*****
/***** SET UP SCREEN *****/
set_up_screen()
{
int length;
div_t x;

BLUE_SCREEN;
CYAN;

```

```

THICK_LINE;
rectangle(50,45,590,300);

outtextxy(40,295,"0");
outtextxy(25,45,"255");
GREY;
rectangle(0,460,630,480);
setfillstyle(1,7);
floodfill(10,470,7);
WHITE;
length = strlen(file_name);
x = div(length,2);
outtextxy((315 - (x.quot*8)),360,file_name);
RED;
outtextxy(5,467,"Program: Analyse.");
outtextxy(365,467,"Written by R.D.F. Keith June 1990.");
}
/*****
/***** Y axis scales *****/
/***** draw colour = grey *****/
/**** draw box for program info ****/
/**** set solid grey fill style **/
/**** fill in box area in grey ****/
/**** draw colour = white *****/
/**** find number characters *****/
/**** divide by 2 (centre text) **/
/***** draw colour = red *****/
/*****
/***** DISPLAY SCALES *****/
display_scales(int i, int zoom_area)
{
int length;
CYAN;
delete_line(310);
if(i == 30)
{
hrs = hours;
mins = minutes;
secs = seconds;
join_strings(hrs, mins, secs);
outtextxy(50,310,scale);

secs = (seconds + 30);
time_check();
join_strings(hrs, mins, secs);
length = strlen(scale);
outtextxy(590 - (length * 8),310,scale);
}

if(i == 6)
{
/* scaling for data 6 secs long */
secs = seconds + ((zoom_area - 1) * 6); /* time of section start */
mins = minutes;
hrs = hours;
time_check();
join_strings(hrs, mins, secs);
outtextxy(50,310,scale);

secs = secs + 6;
time_check();
join_strings(hrs, mins, secs);
length = strlen(scale);
outtextxy(590 - (length * 8),310,scale);
}
}
/*****
/***** TIME CHECK *****/
/* Convert scale times to base 60 */

time_check()
{
if(secs > 59)
{
secs = (secs - 60);
mins = (mins + 1);
if(mins > 59)
{
mins = (mins - 60);
hrs = (hrs + 1);
}
}
}
/*****

```

```

/***** JOIN STRINGS *****/
/* Joins hour$, minute$ and second$ together to form the axis scale string */
join_strings(int hrs, int mins, int secs)
{
char hour[5], minute[20], second[20];

itoa(hrs, hour, 10); /* convert int. (base 10)*/
itoa(mins, minute, 10); /* to a string *****/
itoa(secs, second, 10);
strcpy(scale, hour); /* copy hour$ to scale$ *****/
strcat(scale, ":"); /* append scale$ with ':' *****/
strcat(scale, minute); /* append scale$ with minute$ *****/
strcat(scale, ":"); /* append scale$ with ':' *****/
strcat(scale, second); /* append scale $ with second$ *****/
}
/*****/

/***** DRAW *****/
draw(int section, int step)
{
int x_cord, y_cord;

WHITE; /*** draw colour = white *****/
SOLID_LINE; /*** solid line thickness = 1 *****/

for(x_cord = 1; x_cord < 540; x_cord++) /*** for each x coordinate *****/
{
y_cord = 300 - data[((section - 1)*3000) + (step*x_cord)];

/*** NOTE *****/
/* 300 = offset, section = area for zoom; =1 for no zoom, no. points per **/
/* zoom section = 3000 (6 secs), step = 28 for 30 sec display or = 4 for **/
/* 6 sec display */
/*** *****/
lineto((x_cord+50), y_cord); /* draw from CP to new coord *****/
}
}
/*****/

/***** DELETE LINE *****/
delete_line(int line_no)
{
int colour, i;

colour = getcolor(); /* find out previous draw colour */
LT_BLUE; /* draw colour = light blue *****/
setfillstyle(0, 9); /* set solid fill in light blue **/

rectangle(0, line_no, 640, line_no+8); /* draw rect. over line to del. **/
/* 8 = char height in pixels *****/
floodfill(2, line_no+1, 9); /* fill the line with light blue */
setcolor(colour); /* return to original draw colour*/
}
/*****/

/***** SELECT OPTION *****/
select_option()
{
int i;

YELLOW; /*** draw colour = yellow *****/
outtextxy(50, 400, "select section for zoom");
DASHED_LINE; /*** dashed line thickness = 3 *****/
line(50+108, 45, 50+108, 300);
line(50+216, 45, 50+216, 300);
line(50+324, 45, 50+324, 300); /*** section partitions *****/
line(50+432, 45, 50+432, 300); /*** solid line *****/
SOLID_LINE;
outtextxy(104, 35, "1");
outtextxy((104+108), 35, "2");
outtextxy((104+216), 35, "3");
outtextxy((104+324), 35, "4"); /*** print section numbers *****/
outtextxy((104+432), 35, "5");
}

```

```

/*****
/***** OPTIONS *****/
/***** options menu *****/

options()
{
char ch;
printf("\n\tSelect option:\n");
printf("\n\tZoom data");
INPUTCOLOUR;
printf("\t (z)\n");
SETCOLOUR;
printf("\n\tScroll data");
INPUTCOLOUR;
printf("\t (s)\n");
SETCOLOUR;
printf("\n\tTime jump");
INPUTCOLOUR;
printf("\t (t)\n");
SETCOLOUR;
printf("\n\tAnalyse data");
INPUTCOLOUR;
printf("\t (a)\n");
SETCOLOUR;
printf("\n\tChange file");
INPUTCOLOUR;
printf("\t (c)\n");
SETCOLOUR;
printf("\n\tView samples");
INPUTCOLOUR;
printf("\t (v)\n");

SETCOLOUR;
printf("\n\tExit Program");
INPUTCOLOUR;
printf("\t (e)\n");
printf("\n");
printf("\tOption:");
INPUTCOLOUR;
printf("\t ");
ch = getch();
SETCOLOUR;
CLS;
if(ch == 's')                               /** scroll option sub-menu *****/
    {
    INPUTCOLOUR;
    printf("\n\t----- SCROLL OPTION -----");
    SETCOLOUR;
    printf("\n\tAt any time during scroll option you can:\n");
    printf("\n\tNext frame");
    INPUTCOLOUR;
    printf("\t(any)\n");
    SETCOLOUR;
    printf("\n\tZoom data");
    INPUTCOLOUR;
    printf("\t(z)\n");
    SETCOLOUR;
    printf("\n\tQuit option");
    INPUTCOLOUR;
    printf("\t(q)\n");
    printf("\t\n");
    printf("\t\nPress any key to Continue");
    getch();
    }

if(ch == 't')                               /** time jump sub-menu *****/
    {
    INPUTCOLOUR;
    CLS;
    printf("\n\t----- TIME JUMP OPTION -----");
    SETCOLOUR;
    printf("\n\tNumber of hours from start of file\n");
    printf("\n\tHours = ");
    INPUTCOLOUR;
    scanf("%d", &hours);

    SETCOLOUR;

```



```

    printf("\n\tNumber of minutes from start of file\n");
    printf("\n\tMinutes = ");
    INPUTCOLOUR;
    scanf("%d", &minutes);
}
return(ch);
}
/*****
/***** SCROLL DATA *****/
scroll_data()
{
int flag;          /* set when quit option taken */
char ch;

flag = 0;
initgraph(&g_driver, &g_mode, "");
while(flag != TRUE)
{
    set_up_scales();
    get_data();          /* get next 30 secs of data */
    if(file_flag == TRUE) /* check for eof */
    {
        return;
    }
    display_data();     /* display data */
    ch = getch();       /* get option */

    if(ch == 'z')
    {
        zoom_data();
    }
    if(ch == 'q')
    {
        flag = TRUE;    /* quit scroll option */
    }
}
cleardevice();       /* clear graphics screen */
}
/*****
/***** SET UP SCALES *****/
set_up_scales()
{
seconds = seconds + 30; /* set secs for start of next 30s */
if(seconds > 59)        /* convert to base 60 */
{
    seconds = seconds - 60;
    minutes = minutes + 1;
}
if(minutes > 59)
{
    minutes = minutes - 60;
    hours = hours + 1;
}
}
/*****
/***** ASSIGN COEFFS *****/
assign_coeffs()
{
int i,k;
float dum;

/* Because of the symmetry associated with the coeffs, it */
/* is only necessary to define half the values. */
/* coeff[0] = coeff[N], coeff[1] = coeff[N-1] etc. */

/***** Coefficients for first stage of decimation *****/

stage1_coeff[0] = 0.0065417;    stage1_coeff[1] = 0.0079029;
stage1_coeff[2] = -0.037461;   stage1_coeff[3] = -0.049154;
stage1_coeff[4] = 0.14579;     stage1_coeff[5] = 0.42665;

/*****

```

```

k = 5;
for(i = 6; i < 12; i++)
    {
        stage1_coeff[i] = stage1_coeff[k];
        k--;
    }
    /* there are 12 coeffs in stage 1 */

/* Same comments apply for coeffs for second stage */

/***** Coefficients for second stage of decimation *****/

stage2_coeff[0] = -2.8901e-4;    stage2_coeff[1] = -5.4028e-4;
stage2_coeff[2] = 0.0007343;    stage2_coeff[3] = 5.9946e-4;
stage2_coeff[4] = -6.7544e-4;   stage2_coeff[5] = -0.0012811;
stage2_coeff[6] = 6.1129e-04;   stage2_coeff[7] = 0.0020902;
stage2_coeff[8] = -1.5175e-4;   stage2_coeff[9] = -0.0030269;
stage2_coeff[10] = -8.3217e-4;  stage2_coeff[11] = 0.0038906;
stage2_coeff[12] = 0.0024733;   stage2_coeff[13] = -0.0043998;
stage2_coeff[14] = -0.0048125;  stage2_coeff[15] = 0.0041845;
stage2_coeff[16] = 0.007763;    stage2_coeff[17] = -0.0028237;
stage2_coeff[18] = -0.011071;   stage2_coeff[19] = -1.1437e-4;
stage2_coeff[20] = 0.014294;    stage2_coeff[21] = 0.0050329;
stage2_coeff[22] = -0.016783;   stage2_coeff[23] = -0.012297;
stage2_coeff[24] = 0.017664;    stage2_coeff[25] = 0.022302;
stage2_coeff[26] = -0.015726;   stage2_coeff[27] = -0.035741;
stage2_coeff[28] = 9.005899e-4; stage2_coeff[29] = 0.054545;
stage2_coeff[30] = 0.0069628;   stage2_coeff[31] = -0.085984;
stage2_coeff[32] = -0.049684;   stage2_coeff[33] = 0.17928;
stage2_coeff[34] = 0.41453;

/*****

k = 34;
for(i = 35; i < 70; i++)
    {
        stage2_coeff[i] = stage2_coeff[k];
        k--;
    }
    /* there are 70 coeffs in stage 2 */
}

/*****

/***** DECIMATE *****/

/** The decimation factor is 4 and the number of stages is 2. *****/
/** function reduces the sampling frequency from 500 to 125 Hz and so *****/
/** reduces the max freq. of interest to 62.5 Hz. This decimated data will */
/** will be used to calculate the power of the mains interference and the **/
/** power associated with the base-line drift *****/

decimate()
{
float x1[20], ytemp;

int decimation_factor_for_1, decimation_factor_for_2, z, j, i, k,
    coeffs_in_1, coeffs_in_2;
int count1, count2;

coeffs_in_1 = 11;    /* N - 1 coeffs in stage 1 *****/
coeffs_in_2 = 69;   /* N - 1 coeffs in stage 2 *****/
decimation_factor_for_1 = 2;
decimation_factor_for_2 = 2;
count1 = decimation_factor_for_1;
count2 = decimation_factor_for_2;
i = 0;
z = 0;

for(j = 0; j < 20; j++)
    {
        x1[j] = 0;
    }
for(j = 0; j < 80; j++)
    {
        x2[j] = 0;
    }
}

```

```

while(i != size)
{
    /* exit when data decimated*/

    for(j = 1; j <= coeffs_in_1; j++) /* shuffle data down in stage1 ***/
    {
        /* array and insert latest value */
        k = coeffs_in_1 - j + 1; /* in x[0]. The oldest value ***/
        x1[k] = x1[k-1]; /* falls off the end *****/
    }
    x1[0] = data[i];

    count1--;
    if(count1 != 0) /* get next point and repeat ***/
    {
        i++;
        continue;
    }

    ytemp = 0;

    /* multiply each array location and add result *****/
    /* to previous result for all elements in stage1 array. */
    for(j = 0; j <= coeffs_in_1; j++)
    {
        ytemp = ytemp + (stage1_coeff[j] * x1[j]);
    }
    /* *****/

    count1 = decimation_factor_for_1;

    for(j = 1; j <= coeffs_in_2; j++) /* shuffle for stage2 array ***/
    {
        k = coeffs_in_2 - j + 1;
        x2[k] = x2[k-1];
    }
    x2[0] = ytemp; /* insert newest value *****/

    ytemp = 0;
    count2--;

    if(count2 != 0) /* get next data point *****/
    {
        i++;
        continue;
    }

    count2 = decimation_factor_for_2;

    for(j = 0; j <= coeffs_in_2; j++) /* multiply each array location */
    {
        /* and add to previous result ***/
        ytemp = ytemp + (stage2_coeff[j] * x2[j]);
    }

    decimated_data[z] = ytemp; /* decimated data point *****/
    /*printf("%f\n",decimated_data[z]);*/
    i++;
    z++; /* increment array offsets *****/
    continue;
}
}
/* *****/

/* *****/ ASSIGN FILTER COEFFS *****/
filter_coeffs()
{
    int k, i;
    /* *****/ 50 Hz extraction filter coefficients *****/

    mains_coeff[0] = -0.0090713;    mains_coeff[1] = -0.0071267;
    mains_coeff[2] = 0.0076893;    mains_coeff[3] = -0.0078048;
    mains_coeff[4] = 0.0061224;    mains_coeff[5] = -0.002653;
    mains_coeff[6] = -0.00133;     mains_coeff[7] = 0.0040765;
    mains_coeff[8] = -0.0044633;   mains_coeff[9] = 0.0027741;
    mains_coeff[10] = -5.9713e-4;  mains_coeff[11] = -1.4045e-4;
    mains_coeff[12] = -0.0014514;  mains_coeff[13] = 0.0043484;
    mains_coeff[14] = -0.0059049;  mains_coeff[15] = 0.0033575;
}

```

```

mains_coeff[16] = 0.0041591;      mains_coeff[17] = -0.014311;
mains_coeff[18] = 0.021959;      mains_coeff[19] = -0.021438;
mains_coeff[20] = 9.793199e-3;   mains_coeff[21] = 0.010739;
mains_coeff[22] = -0.032445;    mains_coeff[23] = 0.045108;
mains_coeff[24] = -0.04051;     mains_coeff[25] = 0.017113;
mains_coeff[26] = 0.017972;     mains_coeff[27] = -0.051012;
mains_coeff[28] = 0.0672;       mains_coeff[29] = -0.057417;
mains_coeff[30] = 0.023052;     mains_coeff[31] = 0.023563;
mains_coeff[32] = -0.06384;     mains_coeff[33] = 0.080726;
mains_coeff[34] = -0.066283;    mains_coeff[35] = 0.0255505;

/*****/
k = 35;
for(i = 36; i < 72; i++)          /* 2nd half of coeff assignment */
{
  mains_coeff[i] = mains_coeff[k];
  k--;                             /* 72 coeffs in 50hz_coeff */
}

/***** Baseline extraction coefficients *****/

baseline_coeff[0] = -0.025121;    baseline_coeff[1] = 0.00399952;
baseline_coeff[2] = 0.0038247;    baseline_coeff[3] = 0.0037413;
baseline_coeff[4] = 0.003744;     baseline_coeff[5] = 0.0038122;
baseline_coeff[6] = 0.0039513;    baseline_coeff[7] = 0.0041396;
baseline_coeff[8] = 0.0043829;    baseline_coeff[9] = 0.0046684;
baseline_coeff[10] = 0.0049972;   baseline_coeff[11] = 0.0053575;
baseline_coeff[12] = 0.0057547;   baseline_coeff[13] = 0.0061706;
baseline_coeff[14] = 0.0066223;   baseline_coeff[15] = 0.0070841;
baseline_coeff[16] = 0.0075798;   baseline_coeff[17] = 0.008091;
baseline_coeff[18] = 8.567101e-3; baseline_coeff[19] = 9.146199e-3;
baseline_coeff[20] = 0.0096813;   baseline_coeff[21] = 0.010214;
baseline_coeff[22] = 0.010749;    baseline_coeff[23] = 0.011297;
baseline_coeff[24] = 0.011854;    baseline_coeff[25] = 0.012407;
baseline_coeff[26] = 0.012953;    baseline_coeff[27] = 0.013491;
baseline_coeff[28] = 0.014014;    baseline_coeff[29] = 0.014525;
baseline_coeff[30] = 0.015022;    baseline_coeff[31] = 0.015502;
baseline_coeff[32] = 0.015971;    baseline_coeff[33] = 0.016417;
baseline_coeff[34] = 0.016849;    baseline_coeff[35] = 0.017258;
baseline_coeff[36] = 0.017627;    baseline_coeff[37] = 0.017992;
baseline_coeff[38] = 0.01832;     baseline_coeff[39] = 0.018616;
baseline_coeff[40] = 0.018882;    baseline_coeff[41] = 0.019117;
baseline_coeff[42] = 0.01933;     baseline_coeff[43] = 0.019503;
baseline_coeff[44] = 0.019639;    baseline_coeff[45] = 0.019743;
baseline_coeff[46] = 0.019808;    baseline_coeff[47] = 0.019842;

/*****/

k = 47;
for(i = 48; i < 96; i++)          /* 2nd half of coeff assignment */
{
  baseline_coeff[i] = baseline_coeff[k];
  k--;                             /* 96 coeffs in baseline_coeff */
}

/***** R wave extraction coefficients *****/

rwave_coeff[0] = 0.0026578;       rwave_coeff[1] = -0.0075544;
rwave_coeff[2] = 0.0273;         rwave_coeff[3] = 0.0097524;
rwave_coeff[4] = 0.0027689;      rwave_coeff[5] = 0.013144;
rwave_coeff[6] = 9.3757e-4;      rwave_coeff[7] = 0.0075557;
rwave_coeff[8] = 0.012592;       rwave_coeff[9] = -0.0024583;
rwave_coeff[10] = 0.012202;      rwave_coeff[11] = 0.0060547;
rwave_coeff[12] = -0.0046728;    rwave_coeff[13] = 0.01384;
rwave_coeff[14] = -0.0054332;    rwave_coeff[15] = -0.004492;
rwave_coeff[16] = 9.418799e-3;   rwave_coeff[17] = -0.019357;
rwave_coeff[18] = -0.0020781;    rwave_coeff[19] = -0.0031937;
rwave_coeff[20] = -0.031494;     rwave_coeff[21] = 0.0005579;
rwave_coeff[22] = -0.024179;     rwave_coeff[23] = -0.037034;
rwave_coeff[24] = -5.707e-5;     rwave_coeff[25] = -0.051089;
rwave_coeff[26] = -0.031424;     rwave_coeff[27] = -8.895001e-3;
rwave_coeff[28] = -0.079171;     rwave_coeff[29] = -9.998999e-3;
rwave_coeff[30] = -0.033802;     rwave_coeff[31] = -0.10254;
rwave_coeff[32] = 0.042121;      rwave_coeff[33] = -0.10817;
rwave_coeff[34] = -0.11582;      rwave_coeff[35] = 0.53442;

/*****/

```

```

k = 35;
for(i = 36; i < 72; i++)
    {
        rwave_coeff[i] = rwave_coeff[k];
        k--;
    }
/* 72 coeffs in rwave_coeff */
}
/*****
/***** EXTRACT MAINS *****/
extract_mains()
{
int i, j, k, z, n, g;
float y;
n = 71;
/* number of coeffs - 1 *****/

if ((mains = fopen("\\mains.dat", "w")) == NULL)
{
printf("cannot open mains.dat");
getch();
}

for(j = 0; j < 100; j++)
{
x2[j] = 0;
}

i = 0;
z = 0;
/* set array offsets = 0 **/

while(i < 3800)
{
for(j = 1; j <= n; j++)
{
k = n - j + 1;
x2[k] = x2[k-1];
}
/* shuffle array *****/

x2[0] = decimated_data[i];
/*printf("%f\n", decimated_data[i]);*/
y = 0;
for(j = 0; j <= n; j++)
{
y = y + (mains_coeff[j] * x2[j]);
}
/* calc new point *****/

temp_array[z] = y;
z++;
if(z == 380)
{
for (g = 0; g < 380; g++)
{
fprintf(mains, "%f\n", temp_array[g]);
}
/* print array to file ****/

z = 0;
}

i++;
}
fclose(mains);
}
/*****
/***** EXTRACT BASE-LINE *****/
extract_baseline()
{
int i, j, k, z, n, g;
float y;
n = 95;
/* number of coeffs - 1 *****/

if ((base = fopen("\\bline.dat", "w")) == NULL)
{
printf("cannot open bline.dat");
getch();
}
/* if can't open bline.dat*/
/* print error message *****/
/* wait for response *****/

for(j = 0; j < 100; j++)
{
x2[j] = 0;
}
/* make coeff array = 0 ****/

```

```

i = 0;
z = 0;
/* set array offsets = 0 */

while(i < 3800)
/* 3800 points *****/
{
    for(j = 1; j <= n; j++)
    {
        k = n - j + 1;
        x2[k] = x2[k-1];
        /* shuffle array *****/
    }
    x2[0] = decimated_data[i];
    /*printf("%f\n",decimated_data[i]);*/
    y = 0;
    for(j = 0; j <= n; j++)
        /* calc new point *****/
        {
            y = y + (baseline_coeff[j] * x2[j]);
        }
    temp_array[z] = y;
    z++;
    if(z == 380)
    {
        for (g = 0; g < 380; g++)
            /* print array to file */
            {
                fprintf(base, "%f\n", temp_array[g]);
            }
        z = 0;
    }
    i++;
}

fclose(base);
}
/*****

/***** EXTRACT R-WAVE *****/
extract_rwave()
{
int i, j, k, z, n, g;
float y;

n = 71;
/* number of coeffs - 1 *****/

if ((rwave = fopen("\\rwave.dat", "w")) == NULL)
{
    CLS;
    /* if can't open mains.dat*/
    printf("cannot open rwave.dat");
    /* print error message *****/

    getch();
    /* wait for response *****/
    return;
}
for(j = 0; j < 100; j++)
/* make coeff array = 0 */
{
    x2[j] = 0;
}

i = 0;
z = 0;
/* set array offsets = 0 */

while(i < 3800)
/* 3800 points *****/
{
    for(j = 1; j <= n; j++)
    {
        k = n - j + 1;
        x2[k] = x2[k-1];
        /* shuffle array *****/
    }
    x2[0] = decimated_data[i];
    /*printf("%f\n",decimated_data[i]);*/
    y = 0;
    for(j = 0; j <= n; j++)
        /* calc new point *****/
        {
            y = y + (rwave_coeff[j] * x2[j]);
        }
    temp_array[z] = y;
    z++;
    if(z == 380)
    {
        for (g = 0; g < 380; g++)
            /* print array to file */
            {

```

```

                fprintf(rwave, "%f\n", temp_array[g]);
            }
            z = 0;
        }
        i++;
    }
fclose(rwave);
}
/*****
/***** CALCULATE MAINS POWER *****/
calculate_mains_power()
{
int i;
float mains_total, mains_mean, squared_total, mean_squared;

mains_total = 0;
squared_total = 0;

if ((mains = fopen("\\mains.dat", "r")) == NULL)
    {
    printf("cannot open mains.dat");          /* if can't open mains.dat*/
    getch();                                  /* print error message *****/
    return;                                    /* wait for response *****/
    }
        /***** LOAD IN MAINS DATA *****/

for(i = 0; i < 3800; i++)
    {
    fscanf(mains, "%f\n", &decimated_data[i]);
    }
        /***** CALCULATE MEAN *****/

for(i = 300; i < 3800; i++)
    {
    mains_total = mains_total + decimated_data[i];
    }

mains_mean = mains_total / 3500;

        /***** SUBTRACT MEAN *****/

for(i = 300; i < 3800; i++)
    {
    decimated_data[i] = decimated_data[i] - mains_mean;
    }
        /***** SQUARE SAMPLES *****/

for(i = 300; i < 3800; i++)
    {
    decimated_data[i] = decimated_data[i] * decimated_data[i];
    }
        /***** CALCULATE MEAN OF SQUARES *****/

for(i = 300; i < 3800; i++)
    {
    squared_total = squared_total + decimated_data[i];
    }
mean_squared = squared_total / 3500;
mains_power = 10 * log10(mean_squared);
fclose(mains);
}
/*****
/***** CALCULATE BASE LINE POWER *****/
calculate_bline_power()
{
int i;
float bline_total, bline_mean, squared_total, mean_squared;
bline_total = 0;
squared_total = 0;
if ((base = fopen("\\bline.dat", "r")) == NULL)
    {
    printf("cannot open bline.dat");          /* if can't open mains.dat*/
    getch();                                  /* print error message *****/
    return;                                    /* wait for response *****/
    }

```

```

    }
    /***** LOAD IN BASE LINE DATA *****/
for(i = 0; i < 3800; i++)
    {
    fscanf(base, "%f\n", &decimated_data[i]);
    }
    /***** CALCULATE MEAN *****/
for(i = 300; i < 3800; i++)
    {
    bline_total = bline_total + decimated_data[i];
    }

bline_mean = bline_total / 3500;
    /***** SUBTRACT MEAN *****/
for(i = 300; i < 3800; i++)
    {
    decimated_data[i] = decimated_data[i] - bline_mean;
    }
    /***** SQUARE SAMPLES *****/
for(i = 300; i < 3800; i++)
    {
    decimated_data[i] = decimated_data[i] * decimated_data[i];
    }
    /***** CALCULATE MEAN OF SQUARES *****/
for(i = 300; i < 3800; i++)
    {
    squared_total = squared_total + decimated_data[i];
    }
mean_squared = squared_total / 3500;
bline_power = 10 * log10(mean_squared);
fclose(base);
}

/*****
/***** CALCULATE R-WAVE POWER *****/
calculate_rwave_power()
{
int i;
float rwave_total, rwave_mean, squared_total, mean_squared;

rwave_total = 0;
squared_total = 0;

if ((rwave = fopen("\\rwave.dat", "r")) == NULL)
    {
    printf("cannot open rwave.dat"); /* if can't open mains.dat*/
    getch(); /* print error message *****/
    return; /* wait for response *****/
    }
    /***** LOAD IN BASE LINE DATA *****/
for(i = 0; i < 3800; i++)
    {
    fscanf(rwave, "%f\n", &decimated_data[i]);
    }
    /***** CALCULATE MEAN *****/
for(i = 300; i < 3800; i++)
    {
    rwave_total = rwave_total + decimated_data[i];
    }
rwave_mean = rwave_total / 3500;
    /***** SUBTRACT MEAN *****/
for(i = 300; i < 3800; i++)
    {
    decimated_data[i] = decimated_data[i] - rwave_mean;
    }
    /***** SQUARE SAMPLES *****/
for(i = 300; i < 3800; i++)

```



```

    {
    decimated_data[i] = decimated_data[i] * decimated_data[i];
    }
    /***** CALCULATE MEAN OF SQUARES *****/

for(i = 300; i < 3800; i++)
    {
    squared_total = squared_total + decimated_data[i];
    }
mean_squared = squared_total / 3500;
rwave_power = 10 * log10(mean_squared);
fclose(rwave);
}
/*****
/***** CALCULATE SIGNAL TO NOISE RATIOS *****/
calculate_sig_to_noise()
{
baseline_snr = rwave_power - bline_power;
mains_snr = rwave_power - mains_power;
}
/*****
/***** SATURATION *****/
saturation()
{
int i;
int saturation_min = 15;
int saturation_max = 240;
int saturation_local;
int p, sat_flag;

num_20sec_blks = 0;
num_5sec_blks = 0;
saturation_local = 0;
saturation_total = 0;
i = 0;
p = 0;

while(i < 15120)
    {
    if(data[i] < saturation_min || data[i] > saturation_max)
        {
        saturation_total++;
        saturation_local++;
        i++;
        continue;
        }
    else
        {
        if(saturation_local < 50)
            {
            i++;
            saturation_local = 0;
            continue;
            }
        if(saturation_local >= 50)
            {
            p = 0;
            while(TRUE)
                {
                i++;
                p++;
                if(i == 15120)
                    {
                    saturation_local = saturation_local + p;
                    saturation_total = saturation_total + p;
                    check_sat_blocks(saturation_local);
                    saturation_local = 0;
                    break;
                    }
                }
            if(data[i] < 20 || data[i] > 220)
                {
                saturation_local = saturation_local + p;
                saturation_total = saturation_total + p;
                break;
                }
            }
        }
    }
}

```

```

        }
        if(p == 250)
        {
            check_sat_blocks(saturation_local);
            saturation_local = 0;
            p = 0;
            break;
        }
    }
    i++;
}
}
}
check_sat_blocks(saturation_local);
saturation_local = 0;
total_num_5sec_blks = total_num_5sec_blks + num_5sec_blks;
total_num_20sec_blks = total_num_20sec_blks + num_20sec_blks;
saturation_so_far = saturation_so_far + saturation_total;
}
/*****/

/***** CHECK SATURATION BLOCKS *****/
check_sat_blocks(int block_size)
{
    if(block_size > 10000)
    {
        num_20sec_blks++;
    }
    if(block_size > 2500 && block_size < 10000)
    {
        num_5sec_blks++;
    }
}
/*****/

/***** DETECT SPIKES *****/
detect_spikes()
{
    int i, p, total, mean, count, spike_count;
    int spike_flag;
    i = 0;
    count = 0;
    total = 0;

    while(count < 126)
    {
        total = 0;
        i = 0 + (count * 120);
        for(p = 0; p < 120; p++)
        {
            total = total + data[i];
            i++;
        }
        mean = total / 120;
        i = 0 + (count * 120);

        for(p = 0; p < 120; p++)
        {
            data[i] = abs(data[i] - mean);
            i++;
        }
        count++;
    }
    i = 0;
    spike_count = 0;
    number_spikes = 0;
    spike_flag = 0;

    while(i < 15120)
    {
        if(data[i] <= 50 )
        {
            i++;
            continue;
        }
        if(data[i] > 50)
        {

```

```

        while(TRUE)
        {
            spike_count++;
            i++;
            if(i == 15120 || data[i] < 50)
            {
                break;
            }
        }
        if(spike_count < 10 && i < 15020)
        {
            for(p = 0; p < 100; p++)
            {
                i++;
                if(data[i] > 50)
                {
                    spike_flag = TRUE;
                }
            }
            if(spike_flag != TRUE)
            {
                number_spikes++;
            }
        }
        spike_count = 0;
        spike_flag = 0;
    }
    i++;
}
spikes_so_far = spikes_so_far + number_spikes;
}
/*****
/***** SIGNAL DROP OUT *****/
signal_drop_out()
{
int i, previous, count;
i = 0;
previous = 0;
drop_out = 0;
count = 0;
while(i < 15120)
{
    previous = data[i];
    i++;
    if(data[i] != previous)
    {
        i++;
        continue;
    }
    while(TRUE)
    {
        count++;
        previous = data[i];
        i++;
        if(data[i] != previous)
        {
            i++;
            break;
        }
    }
    if(count > 100)
    {
        drop_out = count;
    }

    count = 0;
}
drop_out_so_far = drop_out_so_far + drop_out;
}
/*****
/***** TOTALS TO FILE *****/
totals_to_file()
{
fprintf(totals, "\n");
fprintf(totals, "Results for %s\n", file_name);

```

```

fprintf(totals, "\n");
time = time / (18.2 * 60);
if(time < 0)
{
    time = time + 1440;
}
fprintf(totals, "Program execution time = %.0f minutes\n", time);
saturation_so_far = saturation_so_far / 500;
fprintf(totals, "total saturation = %.0f seconds\n", saturation_so_far);
fprintf(totals, "number 20 second blocks = %d\n",
    total_num_20sec_blks);
fprintf(totals, "number 5 second blocks = %d\n",
    total_num_5sec_blks);
drop_out_so_far = drop_out_so_far / 500;
fprintf(totals, "total drop out = %.0f seconds\n",
    drop_out_so_far);
fprintf(totals, "total number of spikes = %d\n",
    spikes_so_far);
fprintf(totals, "number baseline power levels less than 10 dB = %d\n",
    power10);
fprintf(totals, "number baseline power levels between 10 and 20 dB = %d\n",
    power20);
fprintf(totals, "number baseline power levels between 20 and 30 dB = %d\n",
    power30);
fprintf(totals, "number baseline power levels greater than 30 dB = %d\n",
    power40);
}
/*****

/***** OPEN FILES *****/
open_files()
{
    power = fopen("\\power.rdf", "w");
    spike = fopen("\\spike.rdf", "w");
    drop_sat = fopen("\\dropsat.rdf", "w");
    snr = fopen("\\snr.rdf", "w");
}
/*****

/***** RESULTS TO FILE *****/
results_to_file(float x_axis)
{
    fprintf(power, "%3.1f\t%2.2f\t%2.2f\t%2.2f\n", x_axis,
        bline_power, mains_power, rwave_power);
    fprintf(snr, "%3.1f\t%2.2f\t%2.2f\n", x_axis, baseline_snr, mains_snr);
    fprintf(spike, "%3.1f\t%d\n", x_axis, number_spikes);
    fprintf(drop_sat, "%3.1f\t%d\t%d\n", x_axis,
        drop_out, saturation_total);
}
/*****

```

Appendix B

C- software implementation of a three layer backpropagation neural network.

/*

NEURAL NETWORK

*/

/* ROBERT D. F. KEITH JUNE 1991 */

```
// Implementation of a 3 layer, fully connected, feed-forward,
// Backpropagation Neural Network which utilises speeding up algorithms
```

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <string.h>
#include <conio.h>

#define TRUE 1
#define FALSE 0
#define INPUTS 4
#define HIDDEN_LAYER 10
#define OUTPUTS 1

#define MOMENTUM 0.7
#define CLOSE_THRESHOLD 0.04
#define LOOP_MAX 1000000
#define CLS clrscr()
void load();

int h,i,j;

float input[INPUTS];
float desired_output[OUTPUTS];
float error_max=0, error_E=0, perc_close=0, number_close = 0;
float magic_alpha = 0.75;

FILE *fp;
FILE *rms_file;
FILE *in_file;
/***** EPOCH *****/
float epoch = 456;
long file_position[456];
char option, file_path[100];
int number, num_less_01=0;
int first_time_flag = TRUE, dir_flag = FALSE, input_flag = FALSE;
float outputs_close = 0; /* to assess how many outputs are close to desired*/
float perc_outputs_close;
int file_open_flag = FALSE; /* for load() */
int ran_flag = FALSE;
int z;
int ran_num=0;
float rate = 0.3;

main()
{
    float weights_v[INPUTS][HIDDEN_LAYER];
    float prev_weights_v[INPUTS][HIDDEN_LAYER];
    float weights_w[HIDDEN_LAYER][OUTPUTS];
    float prev_weights_w[HIDDEN_LAYER][OUTPUTS];
    float v_offset[HIDDEN_LAYER];
    float prev_v_offset[HIDDEN_LAYER];
    float w_offset[OUTPUTS];
    float prev_w_offset[OUTPUTS];
    float middle_node[HIDDEN_LAYER];
    float output_node[OUTPUTS];
    float d[OUTPUTS];
    float e[HIDDEN_LAYER];
    float temp_var;
    float local_error, rms_error, rms_total=0, rms_epoch=0, loop=0;
    int flag, loop_flag, key, x_axis=0;
    unsigned long rms_loop = 0;
    char test;
    int output_tester;
    flag = FALSE;
```

```

/* ++++++ LOAD IN DATA ++++++ */
/***** V WEIGHTS *****/
fp = fopen("v.dat","rt");
for(h=0; h < INPUTS; h++)
{
    for(i=0; i < HIDDEN_LAYER; i++)
    {
        fscanf(fp,"%f",&weights_v[h][i]);
    }
}
fclose(fp);
/***** PREV V WEIGHTS *****/
fp = fopen("pre_v.dat","rt");
for(h=0; h < INPUTS; h++)
{
    for(i=0; i < HIDDEN_LAYER; i++)
    {
        fscanf(fp,"%f",&prev_weights_v[h][i]);
    }
}
fclose(fp);
/***** W WEIGHTS *****/
fp = fopen("w.dat","rt");
for(i=0; i < HIDDEN_LAYER; i++)
{
    for(j=0; j < OUTPUTS; j++)
    {
        fscanf(fp,"%f",&weights_w[i][j]);
    }
}
fclose(fp);
/***** PREV W WEIGHTS *****/
fp = fopen("pre_w.dat","rt");
for(i=0; i < HIDDEN_LAYER; i++)
{
    for(j=0; j < OUTPUTS; j++)
    {
        fscanf(fp,"%f",&prev_weights_w[i][j]);
    }
}
fclose(fp);
/***** V OFFSETS *****/
fp = fopen("v_off.dat","rt");
for(i=0; i < HIDDEN_LAYER; i++)
{
    fscanf(fp,"%f",&v_offset[i]);
}
fclose(fp);
/***** PREV V OFFSETS *****/
fp = fopen("pre_voff.dat","rt");
for(i=0; i < HIDDEN_LAYER; i++)
{
    fscanf(fp,"%f",&prev_v_offset[i]);
}
fclose(fp);
/***** W OFFSETS *****/
fp = fopen("w_off.dat","rt");
for(j=0; j < OUTPUTS; j++)
{
    fscanf(fp,"%f",&w_offset[j]);
}
fclose(fp);
/***** PREV W OFFSETS *****/
fp = fopen("pre_woff.dat","rt");
for(j=0; j < OUTPUTS; j++)
{
    fscanf(fp,"%f",&prev_w_offset[j]);
}
fclose(fp);
/* ++++++ FINISHED LOADING ++++++ */
fcloseall();
/***** PROGRAM START *****/
rms_file = fopen("c:\\rms.dat", "wt");
loop_flag = TRUE;
randomize();
in_file = fopen("c:\\data\\jenny\\ac_ptr.dat", "rt");

```

```

for(j = 0; j < 400; j++)
{
    fscanf(in_file, "%ld ", &file_position[j]);
    // printf("%d \n", file_position[j]);
}
fclose(in_file);
in_file = fopen("c:\\data\\jenny\\train_ac.dat", "rt");
while(loop_flag == TRUE)
{
    load();
    /***** CALC MIDDLE NODE VALUES *****/
    for(h = 0; h < INPUTS; h++)
    {
        for(i = 0; i < HIDDEN_LAYER; i++)
        {
            middle_node[i] = middle_node[i] + (input[h] * weights_v[h][i]);
        }
    }
    for(i = 0; i < HIDDEN_LAYER; i++)
    {
        middle_node[i] = middle_node[i] + v_offset[i];
        if(middle_node[i] >= -15 && middle_node[i] <= 15)
            middle_node[i] = 1.0 / (1 + (exp(-1.0 * middle_node[i])));
        if(middle_node[i] < -15)
            middle_node[i] = 0.000000;
        if(middle_node[i] > 15)
            middle_node[i] = 1.000000;
    }
    /***** CALC OUTPUT NODE VALUES *****/
    for(i = 0; i < HIDDEN_LAYER; i++)
    {
        for(j = 0; j < OUTPUTS; j++)
        {
            output_node[j] = output_node[j] + (middle_node[i] *
weights_w[i][j]);
        }
    }
    for(j = 0; j < OUTPUTS; j++)
    {
        output_node[j] = output_node[j] + w_offset[j];
        if(output_node[j] >= -15 && output_node[j] <= 15)
            output_node[j] = 1.0 / (1 + (exp(-1.0 * output_node[j])));
        if(output_node[j] < -15)
            output_node[j] = 0.000000;
        if(output_node[j] > 15)
            output_node[j] = 1.000000;
    }
    error_max = 0;
    for(j = 0; j < OUTPUTS; j++)
    {
        error_E = (desired_output[j] - output_node[j]) * (desired_output[j]
- output_node[j]);
        if(error_E > error_max)
            error_max = error_E;
    }
    if(error_max < (magic_alpha * magic_alpha))
        number_close++;
    if((error_max >= (magic_alpha * magic_alpha)) || (magic_alpha <= 0.01))
    {
        /***** CALC ERROR IN OUTPUTS d[j]'s *****/
        for(j = 0; j < OUTPUTS; j++)
        {
            d[j] = output_node[j] * (1 - output_node[j]) * (desired_output[j] -
output_node[j]);
        }
        /***** CALC ERROR IN MIDDLE LAYER e[i]'s *****/
        for(i = 0; i < HIDDEN_LAYER; i++)
        {
            local_error = 0;
            for(j = 0; j < OUTPUTS; j++)
            {
                local_error = local_error + (weights_w[i][j] * d[j]);
            }
            e[i] = middle_node[i] * (1 - middle_node[i]) * local_error;
        }
        /***** ADJUST THE W WEIGHTS *****/
        for(i = 0; i < HIDDEN_LAYER; i++)

```



```

{
    for(j = 0; j < OUTPUTS; j++)
    {
        temp_var = weights_w[i][j];
        weights_w[i][j] = weights_w[i][j] + (rate * d[j] * middle_node[i])
            + (MOMENTUM*(weights_w[i][j]-prev_weights_w[i][j]));
        prev_weights_w[i][j] = temp_var;
    }
}

/***** ADJUST THE W OFFSETS *****/
for(j = 0; j < OUTPUTS; j++)
{
    temp_var = w_offset[j];
    w_offset[j] = w_offset[j] + (rate * d[j]) + (MOMENTUM * (w_offset[j] -
        prev_w_offset[j]));
    prev_w_offset[j] = temp_var;
}

/***** ADJUST THE V WEIGHTS *****/
for(h = 0; h < INPUTS; h++)
{
    for(i = 0; i < HIDDEN_LAYER; i++)
    {
        temp_var = weights_v[h][i];
        weights_v[h][i] = weights_v[h][i] + (rate * input[h] * e[i]) +
            (MOMENTUM * (weights_v[h][i] - prev_weights_v[h][i]));
        prev_weights_v[h][i] = temp_var;
    }
}

/***** ADJUST THE V OFFSETS *****/
for(i = 0; i < HIDDEN_LAYER; i++)
{
    temp_var = v_offset[i];
    v_offset[i] = v_offset[i] + (rate * e[i]) + (MOMENTUM * (v_offset[i] -
        prev_v_offset[i]));
    prev_v_offset[i] = temp_var;
}
}

/***** CALC ROOT MEAN SQUAD ERROR AT OP'S *****/
rms_error=0;
for(j = 0; j < OUTPUTS; j++)
{
    rms_error = rms_error + ((desired_output[j] -
        output_node[j])*(desired_output[j] - output_node[j]));
}
rms_error = rms_error/OUTPUTS;
rms_error = sqrt(rms_error);
rms_total = rms_total + rms_error;
rms_error = 0;
/***** CALC TO SEE IF EXAMPLE CLOSE TO DESIRED RESPONSE *****/
output_tester = 0;
for(j = 0; j < OUTPUTS; j++)
    if((desired_output[j] - output_node[j]) * (desired_output[j] -
        output_node[j]) >= CLOSE_THRESHOLD )
        output_tester = 1;
if(output_tester == 0)
    outputs_close++;
if(rms_loop == epoch)
{
    ran_flag = FALSE;
    randomize();
    x_axis++;
    rms_epoch = rms_total/rms_loop;
    perc_close = number_close*100 /rms_loop;
    number_close = 0;
    perc_outputs_close = (outputs_close*100)/rms_loop;
    outputs_close = 0;
    fprintf(rms_file,"%d\t%.4f\n", x_axis, perc_outputs_close);
    if(perc_close >= 91.0 && magic_alpha > 0.01)
    {
        magic_alpha = magic_alpha * 0.75;
        if(magic_alpha < 0.01)
            magic_alpha = 0.01;
    }
    rms_total = 0;
    rms_loop = 0;
}
key = kbhit();

```

```

if(key != 0)
{
    option = getch();
    if(option == 'p')
    {
        if(flag == FALSE)
            flag = TRUE;
        else
            flag = FALSE;
    }
    if(option == 'i')
    {
        if(input_flag == FALSE)
            input_flag = TRUE;
        else
            input_flag = FALSE;
    }
    if(option == 'f')
        printf("%s\n", file_path);

    if(option == 'q')
    {
        loop_flag = FALSE;
    }
}
if(flag == TRUE)
{
    printf("\n");
    //getch();
    for(z = 0; z < OUTPUTS; z++)
        printf("%.0f des = %.0f act = %f rms = %.4f close = %.3f
                alpha = %.3f\n", loop, desired_output[z], output_node[z],
                rms_epoch, perc_outputs_close, magic_alpha);
}
rms_loop++;
loop++;
if(loop - LOOP_MAX >=0)
{
    fcloseall();
    CLS;
    break;
}
}
/***** PROGRAM END *****/

/* ++++++ SAVE DATA ++++++ */
fcloseall();
/***** V WEIGHTS *****/
fp = fopen("v.dat", "wt");
for(h=0; h < INPUTS; h++)
{
    for(i=0; i < HIDDEN_LAYER; i++)
    {
        fprintf(fp, "%f ", weights_v[h][i]);
    }
}
fclose(fp);
/***** PREV V WEIGHTS *****/
fp = fopen("pre_v.dat", "wt");

for(h=0; h < INPUTS; h++)
{
    for(i=0; i < HIDDEN_LAYER; i++)
    {
        fprintf(fp, "%f ", prev_weights_v[h][i]) ;
    }
}
fclose(fp);
/***** W WEIGHTS *****/
fp = fopen("w.dat", "wt");

for(i=0; i < HIDDEN_LAYER; i++)
{
    for(j=0; j < OUTPUTS; j++)
    {
        fprintf(fp, "%f ", weights_w[i][j]) ;
    }
}

```

```

}
fclose(fp);
/***** PREV W WEIGHTS *****/
fp = fopen("pre_w.dat", "wt");

for(i=0; i < HIDDEN_LAYER; i++)
{
    for(j=0; j < OUTPUTS; j++)
    {
        fprintf(fp, "%f ", prev_weights_w[i][j]);
    }
}
fclose(fp);
/***** V OFFSETS *****/
fp = fopen("v_off.dat", "wt");

for(i=0; i < HIDDEN_LAYER; i++)
{
    fprintf(fp, "%f ", v_offset[i]);
}
fclose(fp);
/***** PREV V OFFSETS *****/
fp = fopen("pre_voff.dat", "wt");

for(i=0; i < HIDDEN_LAYER; i++)
{
    fprintf(fp, "%f ", prev_v_offset[i]);
}
fclose(fp);
/***** W OFFSETS *****/
fp = fopen("w_off.dat", "wt");
for(j=0; j < OUTPUTS; j++)
{
    fprintf(fp, "%f ", w_offset[j]);
}
fclose(fp);
/***** PREV W OFFSETS *****/
fp = fopen("pre_woff.dat", "wt");
for(j=0; j < OUTPUTS; j++)
{
    fprintf(fp, "%f ", prev_w_offset[j]);
}
fclose(fp);
/* +++++ FINISHED SAVING +++++ */
fcloseall();
}
/* load for dip classification */
void load()
{
    char file_num_str[15];
    int seek_ptr;
    while(TRUE)
    {
        ran_num = random(400);
        if(ran_num >=0 && ran_num < 400)
            break;
    }
    // printf("%d ", ran_num);
    if(ran_num >= 0 && ran_num < 200)
    {
        desired_output[0] = 1.000;
    }
    if(ran_num >= 200 && ran_num < 400)
    {
        desired_output[0] = 0.000;
    }
    fseek(in_file, file_position[ran_num], 0);
    fscanf(in_file, "%f ", &input[0]);
    fscanf(in_file, "%f ", &input[1]);
    fscanf(in_file, "%f ", &input[2]);
    fscanf(in_file, "%f ", &input[3]);
    if(input_flag == TRUE)
        printf("%f %f %f %f\n", input[0], input[1], input[2], input[3]);
}

```

Appendix C

A transcript from a knowledge elicitation session.

Transcript from knowledge elicitation session 2

24th October 1990

Attending: Engineers
 E Ifeachor,
 R Keith,
 Experts
 J Westgate,
 J Smith.

Key: 1 second pause = .
 2 second pause = .. and so on
 separation - = unfinished statement
 Speaker is indicated by surname initial

Discussion of questions 3 and 4

I Question 3 is er where we wanted you to write down in any convenient form, eg rules, decisions, tree diagrams, whatever, how you actually score the EC - the um heart rate, how do you analyse them, how do you interpret them, you may use diagrams or rule whatever the other point is how often is this done, how regularly do you .. score the ECG or the heart rate pattern really, you have already drawn some diagrams of something

W Yeah, I have, I have drawn some diagrams

I Do you want to talk us through that or its entirely up to you, want to er

W Well . I mean- . I think what I do is look at the thing and just- the first thing that I notice, the baseline heart rate whether its er, and the normal range is sort of like 110 to 160 and that's the first thing that catches my eye, oh you know, where is it situated and then, and then, the next thing I notice is the variability, how much does it squiggle up and down um and that's the first thing that catches my eye, and the I sort of say well are there any accelerations on it and then are there any decelerations on it, and if there are when do they occur and then I look at the contraction channel at that stage to see how frequently the contractions are and then I sort of have a basic opinion in my head as to whether this is normal, ah well probably if I've just been called to look at it, then, as a CTG then I may not if its plum normal, I may just go and ask those other things just to sort of- just to make sure everything is progressing well because there must be some reason they called you, or if I'm in a room just reviewing and- er at hand over time or something, and that's er that's the sort of scheme

I When you say you look at the er um CTG, to see- to look at the baseline and so on, these are usually long pieces of paper running through, (JW yep) do you look at every minute of it or (JW no I usually ah-) 5 minute, what do 60 you do

W Probably um look over at least .. the current 20 to 30 minutes and then I will look back as well, with what Jeremy and I were saying, you always look back to what was happening at the start depending on how long the trace is, so you may sit there and unfold the stuff just like this (visual demonstration with CTG trace) quickly and say yeah yeah that's normal that's normal that's normal and this is normal this is normal here, you've got

a little bit of a problem here but I'm not really worried because this was so normal yes you do have to, if I'm coming to review something as a registrar then I would always look at the whole- I look at the immediate trace and I look at the rest of the trace as well

I What do you look at um, because when I look at some of these waveforms, some of these heart rate patterns, .. I can see a very rapid change over a very short interval in time, does that mean anything? I mean if it doesn't persist for some time does that matter at all? ..

W Well, it depends what- (EI if you suddenly see a dip, a dip there back again or something) a sudden, a sudden 1 deceleration I wouldn't . necessarily be very concerned about, again I would want to look at what's been before and I would want to know if there had been a vaginal examination or she's been sick or been on the bed pan or epidural top up 20 minutes ago, I want to know what kind of dip it was

I Do you want to add anything to this er Jeremy?

S I think exactly the same um . the first thing I look at is as Jenny, that's to determine, there are a couple of patterns which will necessitate instant action which is normally a reflex action but that's rare, so the rest of the time then, I try to relate what I've been called to see to a clinical event exactly as Jenny described ... having got all that information .. if things aren't intending to proceed you then look for, I look for a trend .. and then the trend I relate to the pattern of labour, for example, in a very rapid labour you can sort of get a trend if the trend matches what I would expect then we carry on or do whatever action but again using all the things Jenny said

I Very similar to Jenny (JW yeah JS yes), you talked about the ranges or the range of the baseline, 110 to 160 I believe if it is for example 160 .. all the time-

W If its 160 all the time I will need to say ok what's the rest of it like, ah but essentially a higher baseline would make me want to say um what's the temperature of the 60 of the woman in labour, has she got an epidural in and how long has it been in, because we know that women with an epidural for long time will get hotter .. as they get hotter they get far-, higher pulse rate the baby gets hotter and puts its heart rate up so . um I want to know-, sometimes um high baseline rate is just an indication of a hot baby and a hot mother or sometimes its the beginning sign of um I wouldn't say hypoxic changes but along those lines and so I want to know if there's anything else that makes me worried about this baby, is there meconium ah . ye know um dips, what's the variability like, so I want to know all those things and I want to look back . with the trace to try and decide- .. I think an uncomplicated heart rate of 160 or a 170- . when I mean uncomplicated I mean theres no dips in it or (EI yes) anything, then I'm quite happy for that to continue as it is but if it was what we call a complicated tachycardia with some late decelerations, then I think that's, that's one of the, one of the patterns of the trace where you think, well I think I better perhaps take a closer look at this and if its been persistent then I will want to do a fetal blood sample

I It seems to me that ah regardless of what the heart rate is doing the the decelerations, the dips- are they, are they the warning signs, are those things, when they occur, you keep talk about dips (JW umm) is that-

W Well, I think that, I think heart rate tells a little, the basal heart rate tells you a little bit but I'm not, I don't get overly up tight about basal heart rate unless its at the extremes . the variability I think, I think's important

I It ought to vary, you know is that-

W The um, the actual, how much the beat- what we call the beat to beat, short term variability, the little jitter on the paper that's important but there are a lot of things that affect, that drugs, pethidine will knock that, sometime an epidural will make that go flat, sometimes the babies asleep so it will go flat . and I think decelerations, I don't get upset with early decelerations .. variable decelerations, that depends what they're like and late decelerations, I think you have to take some notice of them if they've been persistent, I don't think you could ignore um 40 minutes of late decelerations

I What would you say was the er bit that-, we've talked about the various things you look at, but I want to explore a little bit more . where you have .. 'cause you could categorise babies or patients into, shall we say three categories . the very normal ones the the obviously abnormal ones (JW mm) and the bits in between which are a 60bit uncertain, what, what would trigger your mind and say this baby is definitely abnormal or something is abnormal ...

W The abnormal baby?

S And this on the CTG is it?

I On the CTG yes, how would you, how would you define that, what, what would you look for in a CTG to make you say ah! that babies abnormal clearly ..

S An overall abnormality is the most important thing rather than any one individual feature, the thing I'd say about dips is dips are one of the few things on a trace that are never good news

I Prolonged dips or-

S Any dip is never really good news where as all other variables, all other features of a trace really relate to good news, dips are the odd man out there, but having said that they very seldom are bad news but that's why dips are singled out .. they're never good news

W For example, on an antenatal trace, on someone not in labour, a dip is definitely bad news, I mean a dip is abnormal(JS certainly within this stable-) in that category, but yeah I'd agree with that 'cos dips I think dips are not goo- although there's a very good reasons for dips

S They may be good news in the sense they may tell you the woman is fully, or a multip of seven but that's, that's very much sort of fine tuning where, where you say good to a dip but that's just to be naughty a bit, by and large a dip should be regarded as not good news but on the other hand it's certainly not always bad news, but it is something that needs assessment . which is why I said I tend to look at a trend because if you've got a situation that may be bad and it seems to be getting more obvious then it probably is

W Yeah I mean, if I just walked into a room and looked at a trace I would go eek! if the heart rate was say 70 (JS mm)

I For how long?

W If it had been 70 for, I think probably longer than 3 . minutes with no variability, so it's a sudden g-dom! like that (JS yes yes) I would go eek!

S Particularly if that was at the end of a trend (JW that's 60 right) if it was an isolated feature within a good trace you would say eek what's the reason, but you wouldn't go eek I have to do something now

W Exactly, if you saw a trace with lots of decelerations and a nasty pattern before had caught your eye and you saw this clunk! (visual, indicating extreme sudden drop in heart rate) then you say you know you must do something and you must quickly find out if the womans fully and you can do forceps right then and there or if it's no where near fully then you go straight down to theatre I mean you must, you must deliver that baby and you've got to move as fast as possible so I think I'd say eek! to that (JS leaves to answer bleep)

W So I'd say eek to that, um for a for a brady- um definat bradycardia you know below 80 with no (EI variability) variability yes um I'd also say eek! I'd better do something, if a very tachycardic trace a heart rate of 170 or or more with with late- with very poor variability and late decelerations so it was almost like a straight line with just tiny shallow late decelerations I would say eek I don't like this I must do something .. a straight line, and its just going up and down horrendously like the baby just can't control its cardiovascular system, if this happened at any stage I would say I don't like that I want to get this baby out um The other pattern that people keep talking about is what we call a sinusoidal pattern where its just a straight line doing a nice little sine wave, usually its within the normal range, .. that happens very infrequently and so you know its er regarded as being pre terminal

K What is, how deep is a drop in baseline before its regarded as a dip

W More than 5 beats, probably more than 10 beats I would say, 5 beats may be very important because when you when you've got these babies who are really struggling they are tachycardic, the ones that I described before, no variability, and sometimes they seem as though they just don't have the capacity to have decelerations so you might just get a very shallow dip it might be 5 beats .

I There could be all sorts of reasons why theres a dip (JW yep) and I guess one might be um the baby is really in trouble obviously (JW yeah) but it could just be the baby has moved or or whatever, what do you think the major causes of dips (JW reasons for dips) yeah, good ones and bad ones if you like, what you consider the-

W Well the early, the early, you know what I mean by, a deceleration which is the mirror image of the contraction 60 now that is supposedly is associated with head going into the pelvis, with forceps, when you put them on you find heart rate goes down, there is a general theory, I don't know if people are absolutely 100% confident about, that early decelerations that can be a result of the head coming down and certainly in someone who is having a rapid delivery these decelerations can be quite big and that's what Jeremy was talking about the head ah the person looks as if they're coming up to fully dilated, they can be quite small or they can be quite large .. by themselves, like I said before, I don't get too uptight about them, recurrent lar- by large I mean more than 60 beats dropped and more than 60 seconds, for that period of time the baby is not getting oxygenated so it will have a cumulative effect, if you've got a baby in good nick at the start of, if you have a multip whose having a tumultuous labour and the baby is whizzing down the pelvis they will have these enourmous dips, no if the baby's fine to start with it will tolerate those periods of

reduced oxygenation and it will cope with that and it will be alright, other babies who are perhaps not happy to start with when they have these, these prolonged- . even if they are early they will have a cumulative effect, a build up in carbon dioxide, leading to respiratory acidosis and eventually a metabolic acidosis, you ask me what are the other reasons for dips (JS returns)

What are, what are the causes of dips?

S The whole premise is that we don't know, have we established that?

W Well I said early deceleration head compression but that's the theory but is that right?

S There is evi- there is concrete evidence that it isn't, at this meeting it emerged that there are two types of type one dips which is interesting, one type of type one dips almost always never do anything but there's another type that's just . an early type two so- . which was interesting, if it to be true doesn't it (JW I'm sure it is) so we lump them all together by looking at them in terms of time which is terribly spurious really if you don't know what the cause is

W So I think its a fair summary to say that little, little early decelerations on an otherwise normal trace (JS its a good sign because it means shes contracting) yeah, you know we wouldn't write home about them at all

I There was just a slight -, there was a relative whose wife went into labour and to cut it short, it had something wrapped around its neck (JW & JS the cord!), presumably if 60 you monitored the heart rate there would be signs of, on the heart rate

S We think there are classical signs and thats an 'm' shape dip (EI that would show that) yes an 'm' shape dip.

I Would there be any deceleration or what (JS yes) the baby is obviously being staved of-

S its not really being staved its just that with a contraction you are occluding the vein first and that produces a reflex acceleration of the heart, you then occlude all the vessels to, to a varying degree so you get a dip of varying degree but it is a very short lived dip, as soon as the contraction starts to ease off, everything else is as normal so the flow rapidly returns and compensates and you can keep that going for ages without the baby de-compensating, because during a contraction theres never any blood flow, if the cord is around its neck you just get a slightly increased lack of blood flow over a short duration and because everything else is normal that's the dip you get, we explain that in pseudo science as reflex tachycardia etc etc but it's a neat explanation for students, the problem with cord around the neck is if it happens to tighten during passage through the pelvis (EI yes that's what I was wondering actually) if it's a multip she can push in 20 minutes and you've got a healthy baby, 20 minutes of hypoxia doesn't do any harm, but if that 20 minutes becomes an hour then you get a very dominant decompensation but you can see that in the trend

W We call these a variable deceleration so what you'll have is variables which are sort of short and well, moderately deep, 40 beats or something and then as that gets worse the baseline will go up to a tachycardia and you will loose your variability at the top in between contractions, flatter, it will become deeper and longer and soon you'll have, if you leave it

you'll have this pattern I told you before and if you leave that, all of a sudden it will go clunk (visual, indicating sudden drop in baseline heart rate) and you get a terminal bradycardia, so that's the pattern of those and it's just a cumulative thing

K On an ominous dip does the response rather slowly come back up?

S That's exactly right ... If you don't get the pre-dip acceleration, and you get a wild overswing that's supposed to be autonomic imbalance which is a sign of decompensation

W The cord doesn't just have to be around the neck, the cord can be round the body or arm or, or just a loop of cords beside the head can give you problems 60

I To the lay man you see the cord around the neck, you get the impression that the baby is about to be strangled

S Yes, its not the neck that counts its the cord, people do think it's the neck that's important

I And its actually the cord

W The cords the important thing Other things that cause dips are just simple things like going on the bedpan (JS yes) or vomiting (JS I've never established why) well I mean I can show you a trace out there (inference, in the randomised clinical trial records) right now, that has got large decelerations with someone on the bed-pan (EI is that so) whether it's just an increase in pressure as they strain and push down-

S It's weird isn't it because you don't get it when you are squatting and pushing as far as I can see, this is a classic phenomenon where a student midwife will call you for a caeser .. but it's a bed pan dip

W And vomiting can do this type of thing

S A healthy baby will shrug it off but if you've got a dodgy baby that can be the last straw, so that's why we were saying trends what have you (JS leaves)

W The other things that can cause dips is if the mother is lying flat (EI on her back) on her back so that her blood pressure drops and drops perfusion to the placenta and drops the babies heart rate, this frequently happens in the first stage with a vaginal examination . um . you can get a deceleration during that and sometimes during a vaginal examination with someone just pushing on a babies head, you know, this head compression bit, sometimes that will caused a deceleration as well so those are, and again whether that's significant depends on what the baby was like before normally you might see a dip almost like that (quick sketch) with a rapid recovery that's fine, that's alright but if you have this (quick sketch) and it's a slow recovery and it takes a while afterwards you think well the baby took a while to recover from that The other thing that causes dips of course is this epidural business which is very important in this day and age (EI that seems to be quite regular) in our experience, theres one paper which gives it an 11% incidence, I don't know what, (JS returns) we were talking about epidural dips and 11% incidence (JS yes) it will be interesting to see what ours is in a way wouldn't it (JS yes) and there's two reasons for an epidural dip, one is that the mothers blood pressure has dropped and the other is the

mother absorbs some of the local anaesthetic that they use, although its not put into blood vessels, and its in her circulation and gets into the fetus and those drugs, like Marcaine, are very potent central depressants and produce a very long deceleration for 3 to 4 minutes with a very slow recovery for upto 40 or 50 minutes after you, and she's had a top up so I think we will just wait and see what happens ..

S Often you get characteristics where the womans blood pressure hasn't dipped and we have a suspicion that it's the maternal response to what would be a falling blood pressure, so may be its the maternal response that causes the dip, so I think often the mother must compensate and so you don't get a change in blood pressure but the fact that she's compensating or shunting is enough to cause some change in uterine flow but its a bit difficult to prove

W Its those ones that I'd say were the Marcaine absorbtion dips

S But sometimes its a bit quicker isn't it?. its like in 10 minutes

W Yeah - (untranslatable but offers some possible reasons for it - emphasis on not sure why)

I I think we shall come back to this abnormal case, it's perhaps the most important in a way ... what would you then consider to be absolutely normal, that which would give you no concern whatsoever

S Presence of acceleration (EI the pesence of would cheer you up ?) yeah - its just unknown to have complications when you have acceleration by which I mean 15 beats per minute for 15 seconds or more as long as you've got a reasonable baseline which we defined before as being 110 to 160 ...

W And no dips

S Well I don't think you even have

W Well, I think that, I think heart rate tells a little, the basal heart rate tells you a little bit but I'm not, I don't get overly up tight about basal heart rate unless its at the extremes . the variability I think, I think's important

I It ought to vary, you know is that-

W The um, the actual, how much the beat- what we call the beat to beat, short term variability, the little jitter on the paper that's important but there are a lot of things that affect, that drugs, pethidine will knock that, sometime an epidural will make that go flat, sometimes the babies asleep so it will go flat . and I think decelerations, I don't get upset with early decelerations .. variable decelerations, that depends what they're like and late decelerations, I think you have to take some notice of them if they've been persistent, I don't think you could ignore um 40 minutes of late decelerations

I What would you say was the er bit that-, we've talked about the various things you look at, but I want to explore, saying the heart rate is going up and up and up isn't it is that correct?

S No that's what I would regard as a trend, an acceleration is a short lived change (EI so we're talking about short lived) yes, it's 15 seconds to 10 minutes so you can put that as absolute

I Ok so that happens without time interval (JS yes)

W Those are the type of things (visual, presence of acceleration on an actual CTG trace) we're talking about, that's a bit

I It goes up and stays there a little

W It may just go up, up and down like that, it may go up and fiddle around a bit and then come down . um the pattern is variable, the basic thing is it usually lasts .. a acceleration can, an acceleration is often associated with fetal movement you see (EI yes)

S This is going to go a long way to shorten this discussion by hours (visual, CTG paper) A change in baseline requires the rate to change by 10 beats per minute for 10 minutes at least .. and that's the difference with acceleration because these are all the questions you get asked but most people, I don't think, will come up with an answer, we haven't dreamed these up (visual reference to CTG paper) we have got these from at least 3 or 4 different books

I I notice in that diagram (visual reference to heart rate pattern on front page of CTG paper) you haven't actually got the contraction waveform, how would that relate to contraction if you imagined a contraction waveform underneath it

S Nothing described here is a contraction related phenomenon, the only established contraction related phenomenon is a dip (JW sometimes-) an acceleration during a contraction is probably the absolute optimum sign in a trace because it's an attack response .. um again by the same token an acceleration matched to a movement is a standard thing so for a dip, you cannot show a dip without a contraction pattern

K How does the 60 rule fit in with this

W If you have a deceleration that lasts less than 60 seconds you know that oxygen and blood flow is not compromised, after 60 seconds you're talking about a more significant phenomenon for the fetus ...

S The effect will be hypoxia, that hypoxia in a normal fetus then you're fine, but if that's a continued insult it gets more hypoxic each time, ... with this, the thing is I know Mr. Greene disagrees with our definition of a variable dip he says that a variable dip is anything that isn't a type 1 or a type 2, which is too great for me, I couldn't accept that .. well . its just the way I think, you've got a particular dip that can be identified as a variable dip and if they're not variable they are type 2 by my definition, again that wouldn't be a constant thing I think if you were to work out ultrasounds that problem wouldn't surface, I don't think it would matter, that distinction (JW no I-), because a variable dip is perhaps one bit of cord or that's the way we like to think about it . so I just mention that he hasn't actually formally accede or agreed with that (visual reference to CTG paper) because I haven't shown him, but we were just talking and he said that

W Dips can only be early or late can't they (JS yes) so I recognise the pattern and if it's a late one I'm not as excited, if it has these deceleration then I'm not going to get up tight about a late deceleration (JS precisely precisely)

break while we view what JS and JW have produced in the way of tree diagram etc,

W My tree has begun to get very complicated and so I haven't quite finished but you will get a few ideas, the other things I've been trying to do is to get some weightings, some risk assessments, so you would assess the risk at the beginning of the labour, antenatal factors with various weightings which people would obviously disagree with and so you've got a risk factor at the beginning which places a woman in a high, medium or low risk and then you would have events happening during labour which like meconium liquial, slow progress, heart rate abnormalities ... the woman would then have an intrapartum risk assessment which will need to be up dated every now and again and so if you have a woman who's high risk actually enters labour who has high risk changes as opposed to someone who's normal, normal everything's been normal and she has 1 dip, I don't know how you build that into a system but that's how we look at it

I It will evolve as we go along, have you got as far as you need to go on that (visual reference to JW risk assessment)

W I haven't gone as far as I'd like but I will keep working on it

Irrelevant conversation talking about hopes for next meeting, phone rings and JW leaves to answer it

Resumption now looking at some CTG traces

W The first baby, shes had an uneventful pregnancy but she is 15 days post term, so in other words she is over 2 weeks overdue, term is anywhere between 38 weeks and 42 weeks as you start to get beyond 42 weeks you know you start to think well ok that puts her into a high risk, she's been induced with a prostin pessary thing which places her into a risk category as well, she then goes into labour this is the first dip on her trace here (visual reference made to trace) the immediate thing that comes to mind, I say, I say, first of all hang on, there appears to be some degree of baseline tachycardia (EI why do you say that) because up here, I suspect that's the baseline I can immediately see there are some decelerations and I can immediately see that the variability is good, it jiggles up and down a lot, so now I say ok, and look at the contractions and notice that she is contracting quite frequently, now I notice these decelerations, here are late decelerations because they start-, they occur after the contraction ok and I can't see .. a definite pattern like that so these are late decelerations . rapidly contracting, now she gets the waters broken or her membranes ruptured and there is thick thick meconium liquial so that's, so we've got a post mature baby who has been induced, with heart rate irregularities which are significant, and thick meconium liquial so she is a very high risk group, so immediately alarm bells start ringing, now what happens here is she, this is during a vaginal examination so that's probably why that's a bigger dip . um they've given some pethidine here and then you've got the same thing you've got a baseline tachycardia of 170 with late decelerations which are probably only 10 beats to 12 beats of deceleration but those are significant

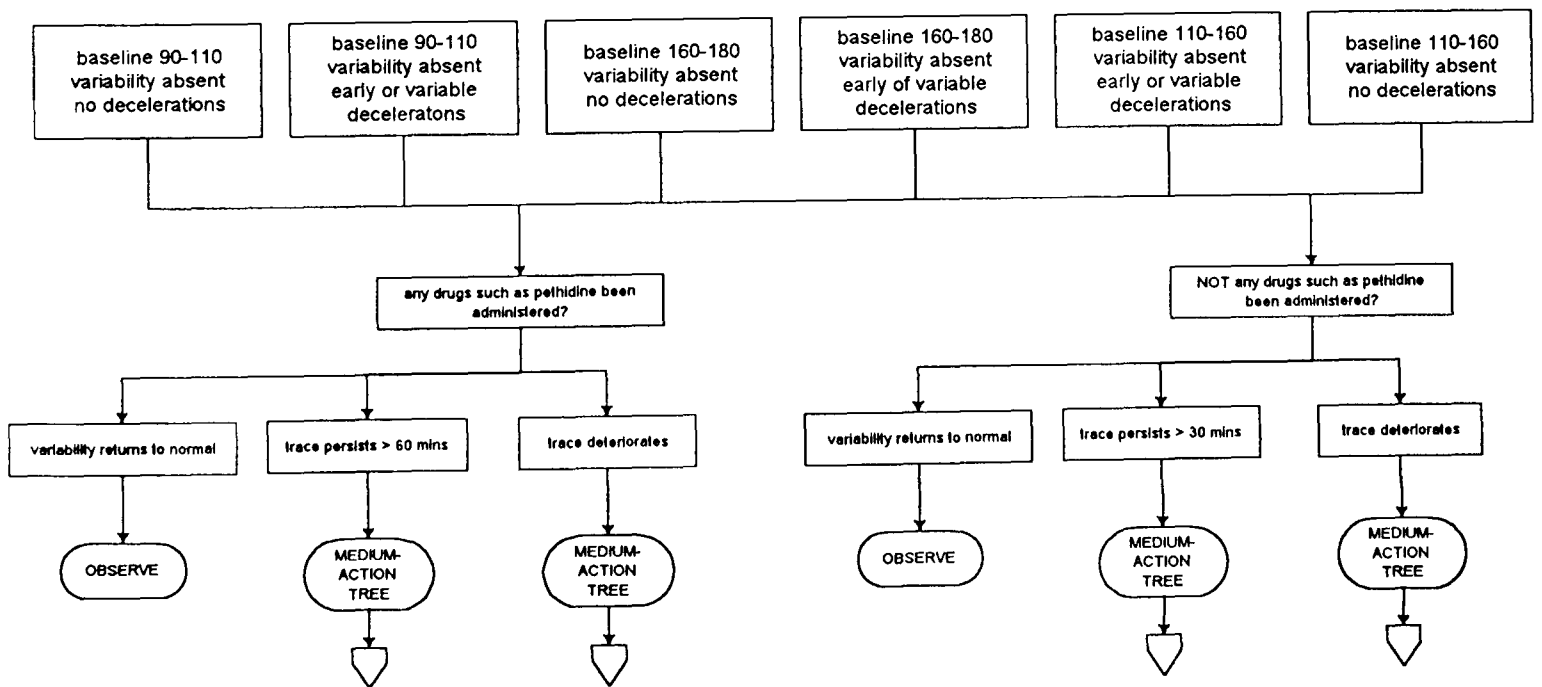
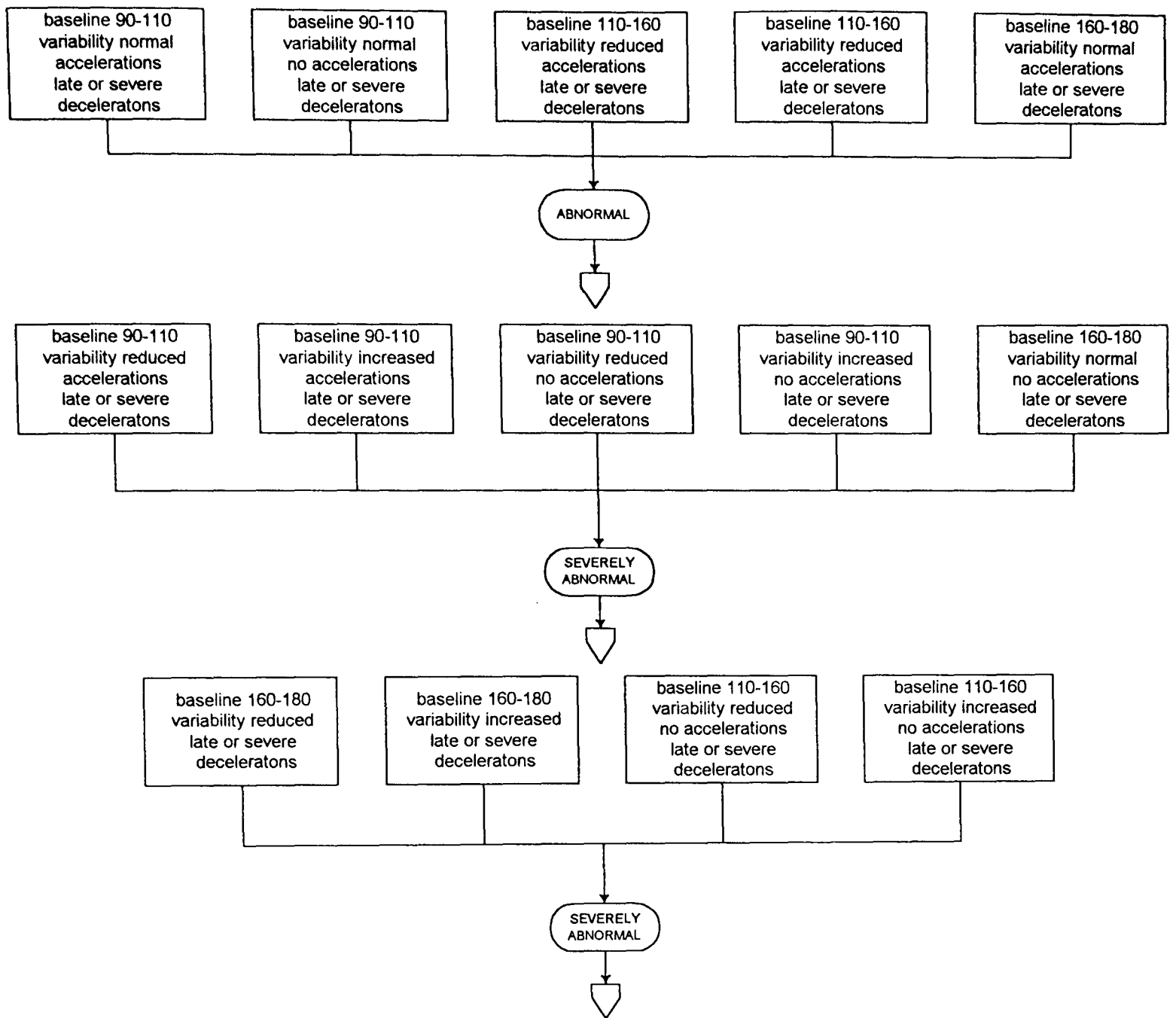
I I want to clear my mind when you say they are significant, we are not overly above 160 which is your magic figure

W Well the baseline up here say is about 170 but it is what I was saying to you, although the variability is reasonable you are getting a late, a dip, a deceleration, you see what I am saying, sometimes these decelerations are not very deep, they are shallow, but that to me, a shallow deceleration on a tachycardic trace are one of the more worrying types, that's significant ok and they're late contraction here, dip there, contraction here dip here, now here we have a bed pan, so there she is sitting up on the bed pan and you get this prolonged . er deceleration which it tries to recover a little and it goes down a bit, and again the dip (EI its gone down quite a lot now) yes it has (EI down to about 100 now) .. now that's something-, the babies had a dramatic change and it's prolonged, so immediately after that I think that's not good news either, this woman already has been on labour ward nearly an hour and I would . have . possibly thought that given all these findings, probably about here (visual reference to CTG trace) the midwife should have called someone and this is a significantly abnormal trace, thick meconium and a post mature baby so she's missed the boat here so it continues and you see that because this doesn't look so bad as that people get fooled and finally she thinks perhaps I'd better talk to someone about this so she gets the SHO ok, so the SHO takes 15 minutes to get there and then the SHO sees the woman and you see these decelerations are now becoming more prolonged um and now the SHO thinks ooo! I better get the registrar, and he comes and thinks oh my goodness! I don't like this at all, if you notice here she's having a long contraction and so she's having quite pronounced decelerations because it's 1, 2, 3 minutes so that's why the heart rate plummets at that point where as the other contractions have been a minute, a so you're only gonna get a shallow deceleration, he then does a fetal blood sample and the ph is 7.15 which is quite abnormal shes only 4 cm so she has an immediate caesarian section (JW leaves to answer phone) so this illustrates nicely a trace which we think is abnormal which was missed, the response time of the midwife is about an hour and a half, now fortunately because this baby has got a good variability here we would expect that although it would come out a bit acidotic I think you could say its not a happy baby but it isn't a damaged baby in fact it was a bit slow to go as we say and it was acidotic but not metabiotically so on its cord gases, it took about 6 hours to recover but it was fine and was back in the ward with its mother, but if you left that longer you probably would have been in trouble.

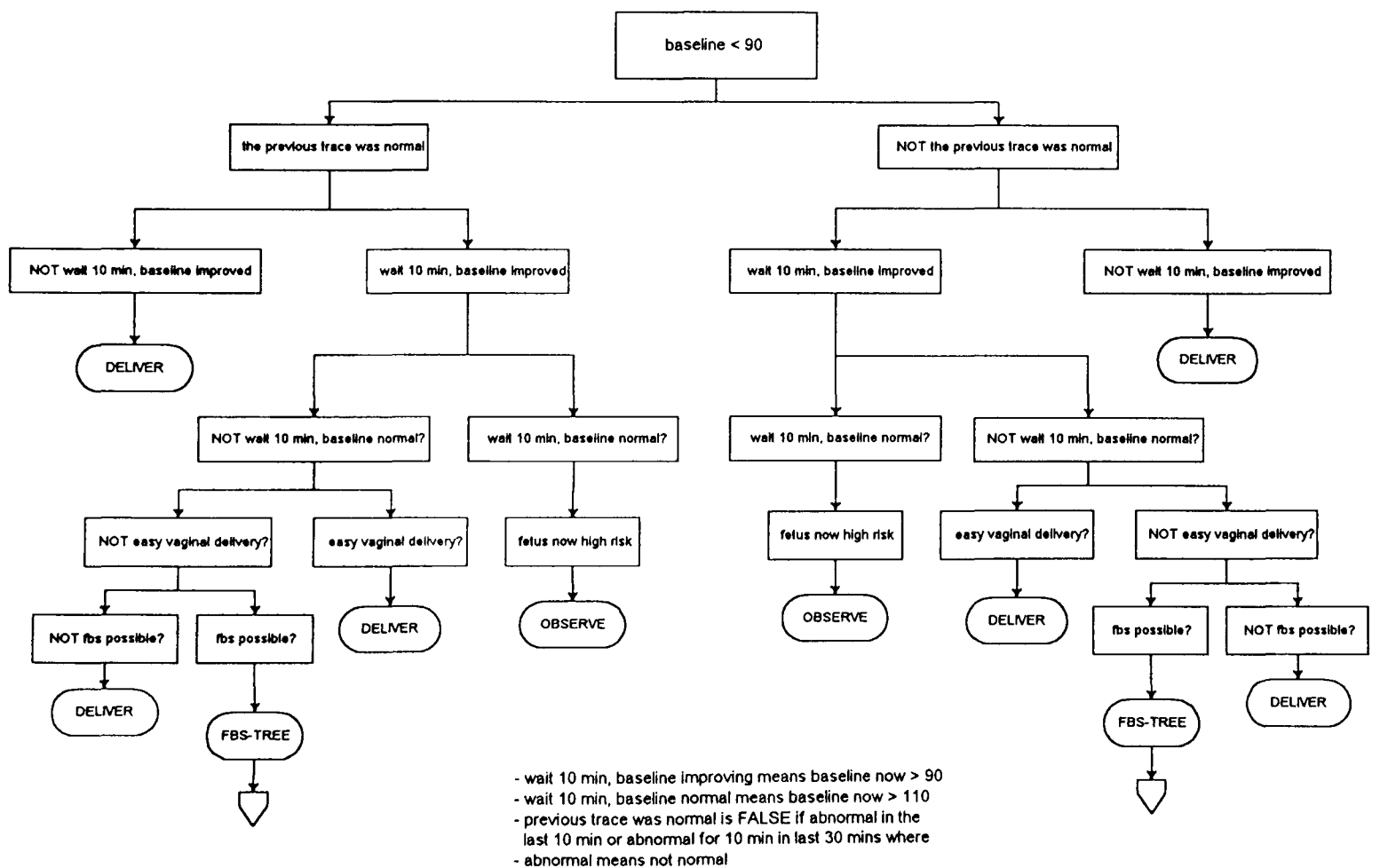
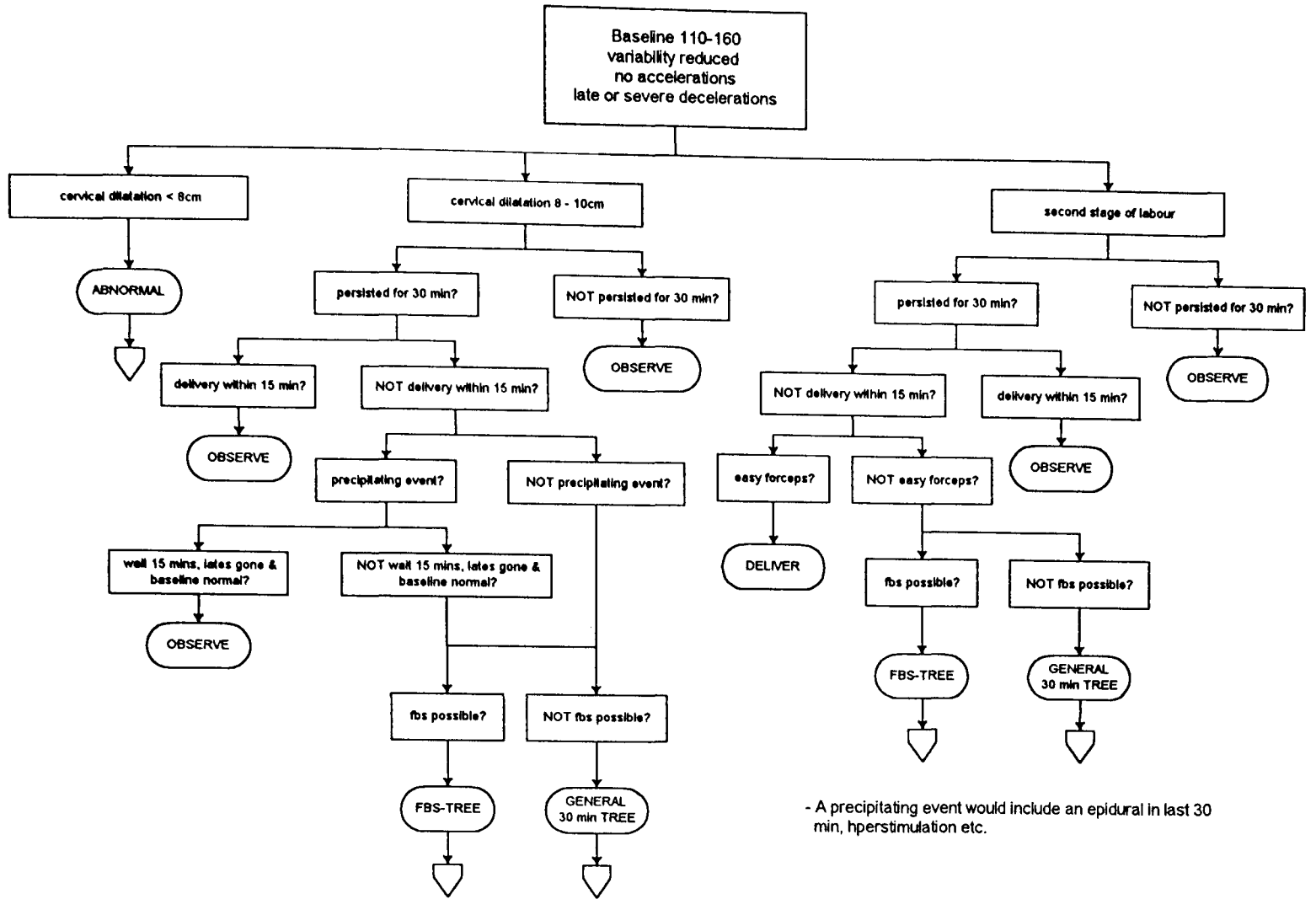
Session closed.

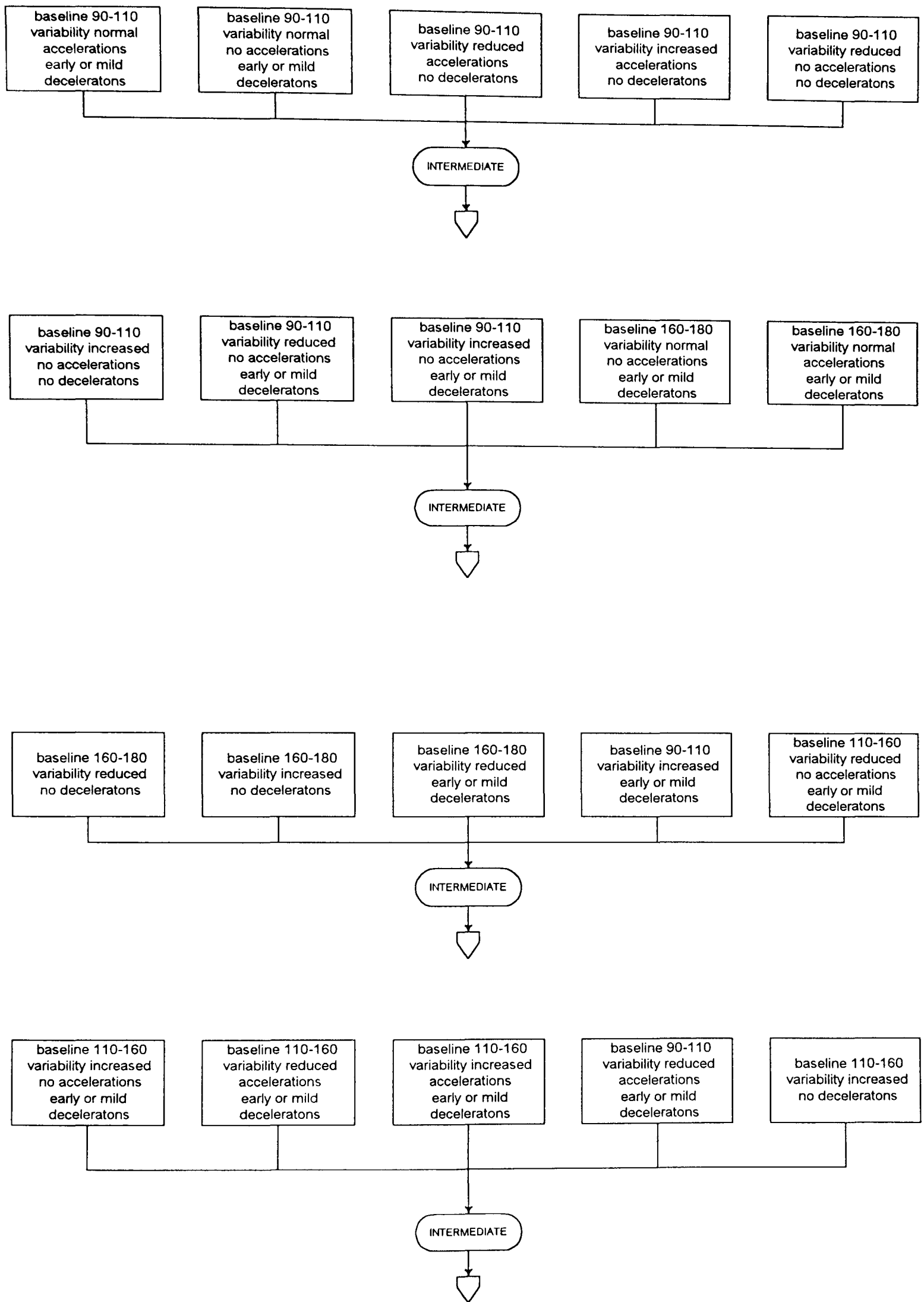
Appendix D

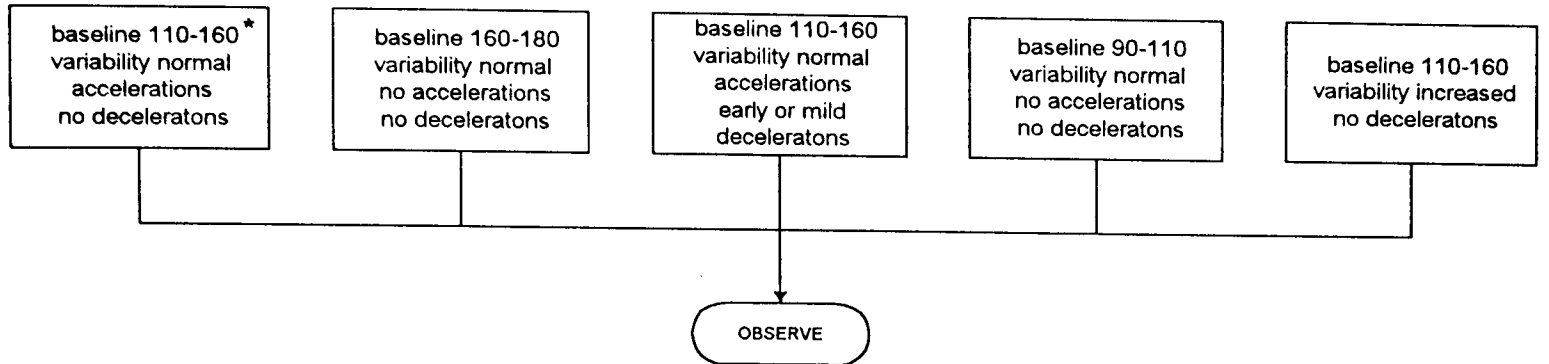
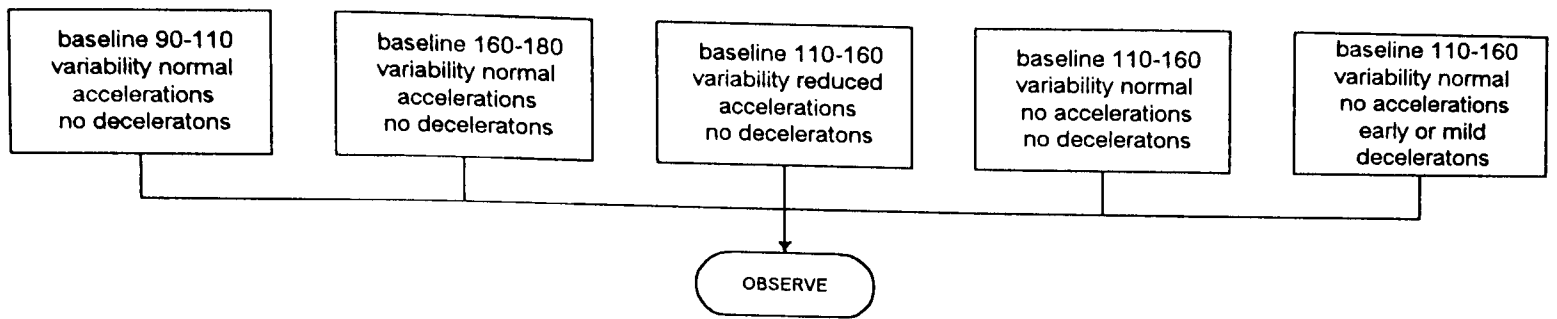
The knowledge tree.



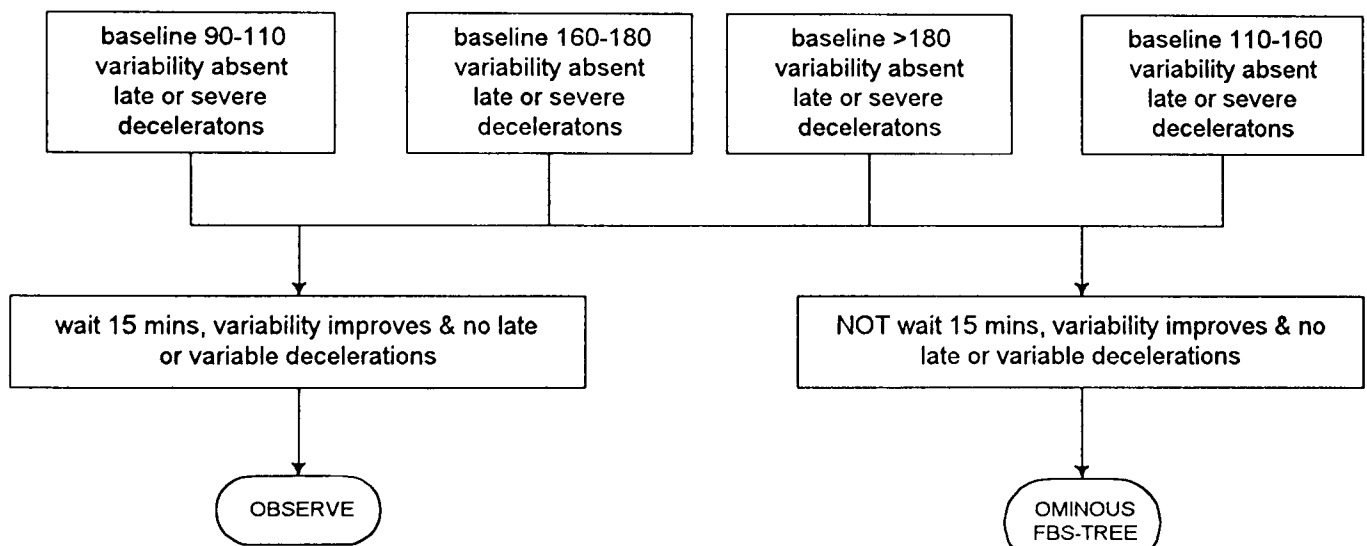
- for normal, the trace must be entirely normal for 10 mins
 - the trace deteriorates IF
 decelerations become late or severe
 OR IF baseline not normal & goes < 90 or > 180
 OR IF baseline was normal and goes 90-110 or 160-180



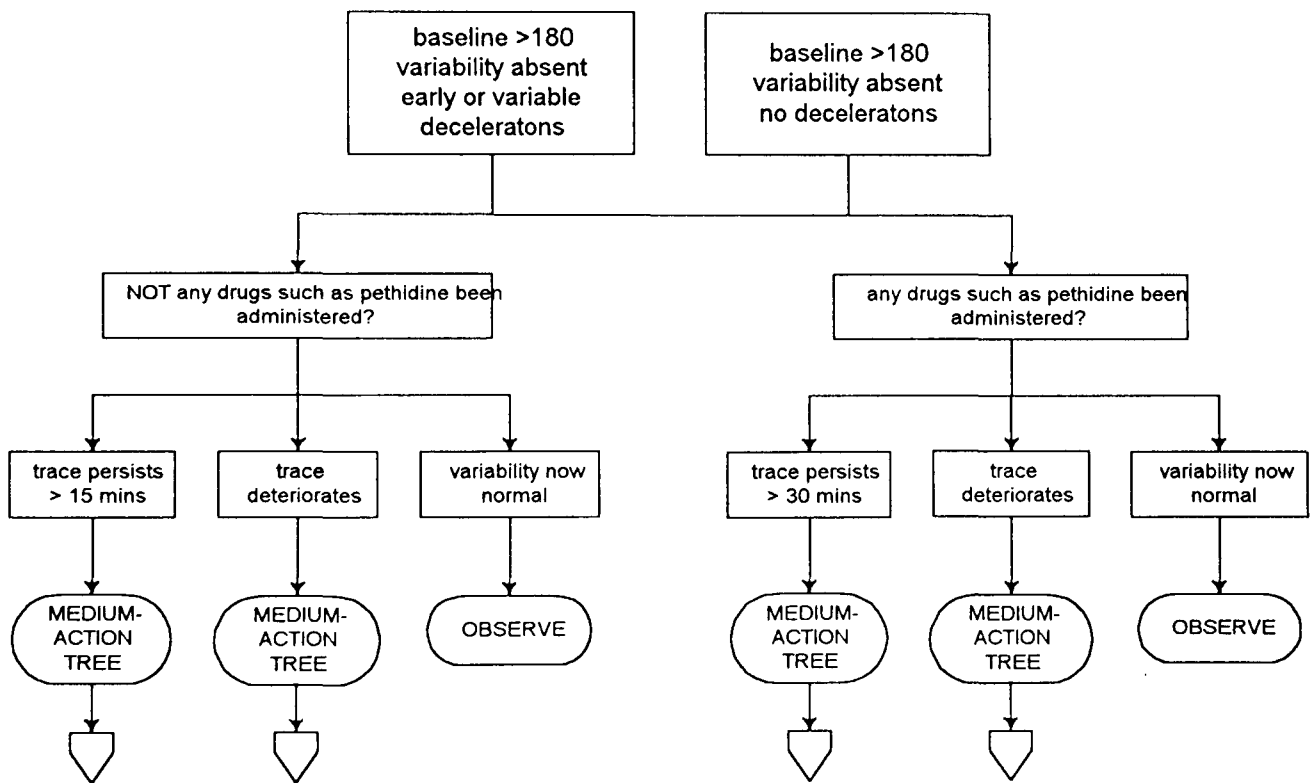
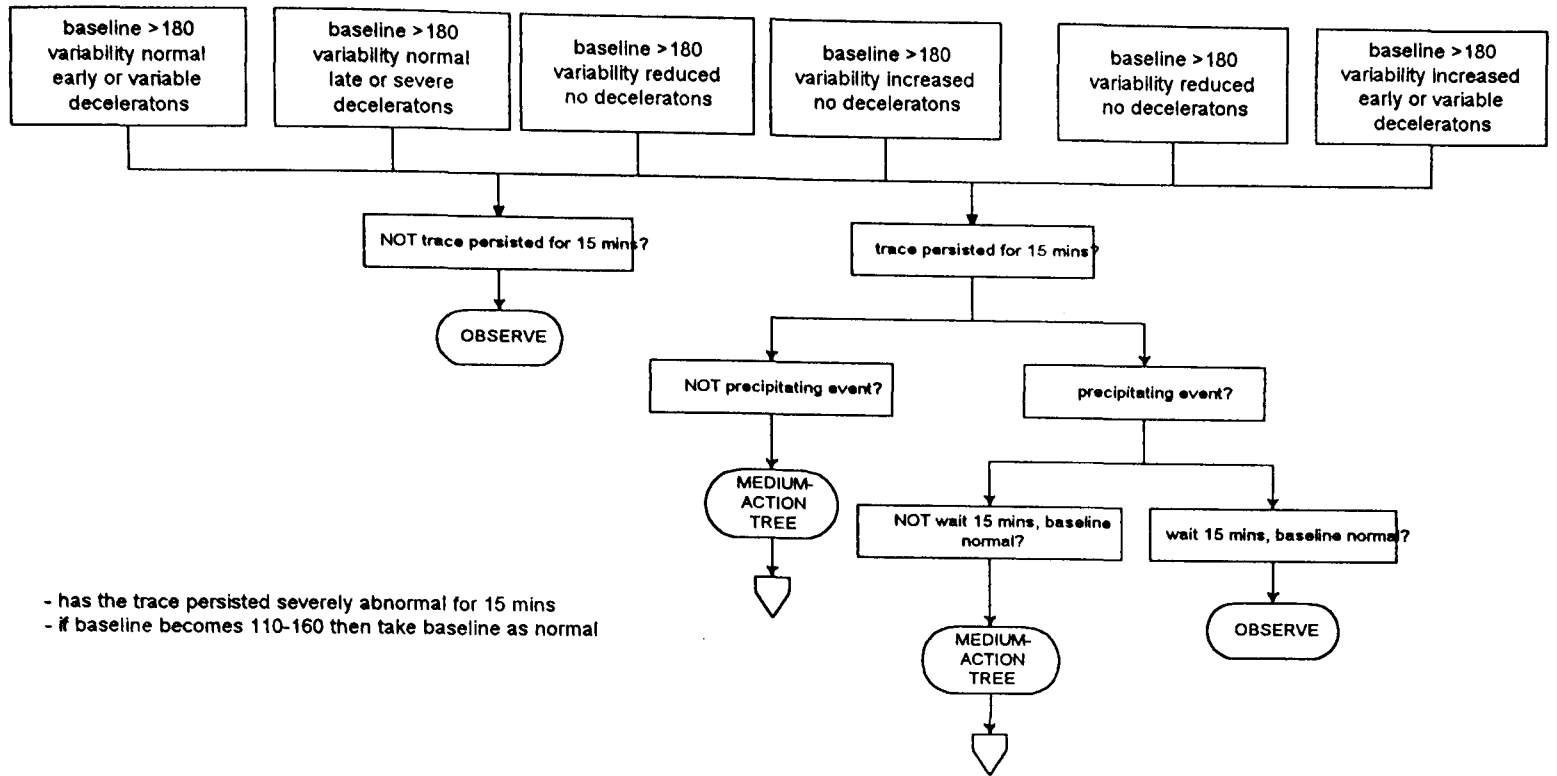




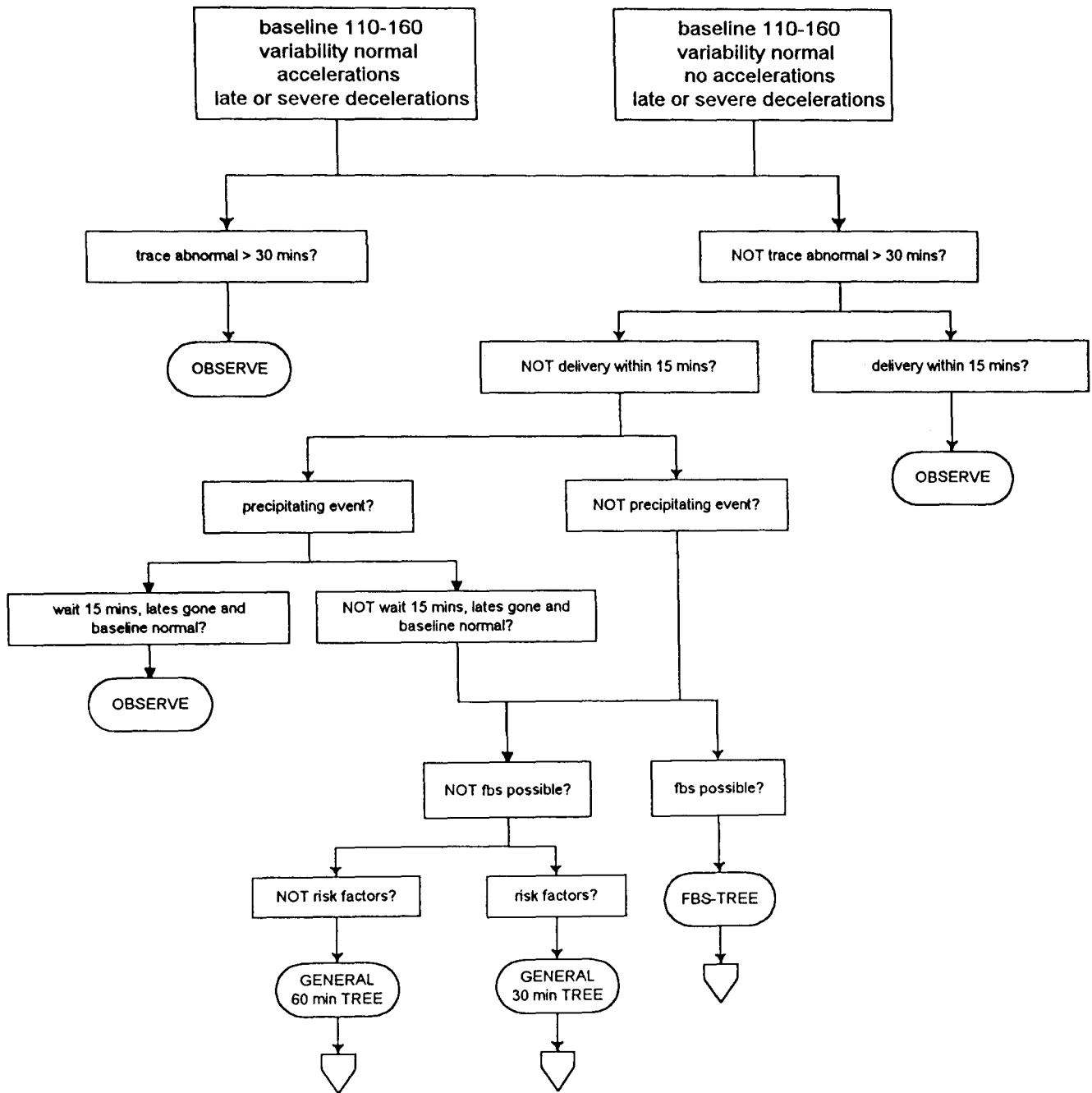
* normal reactive CTG



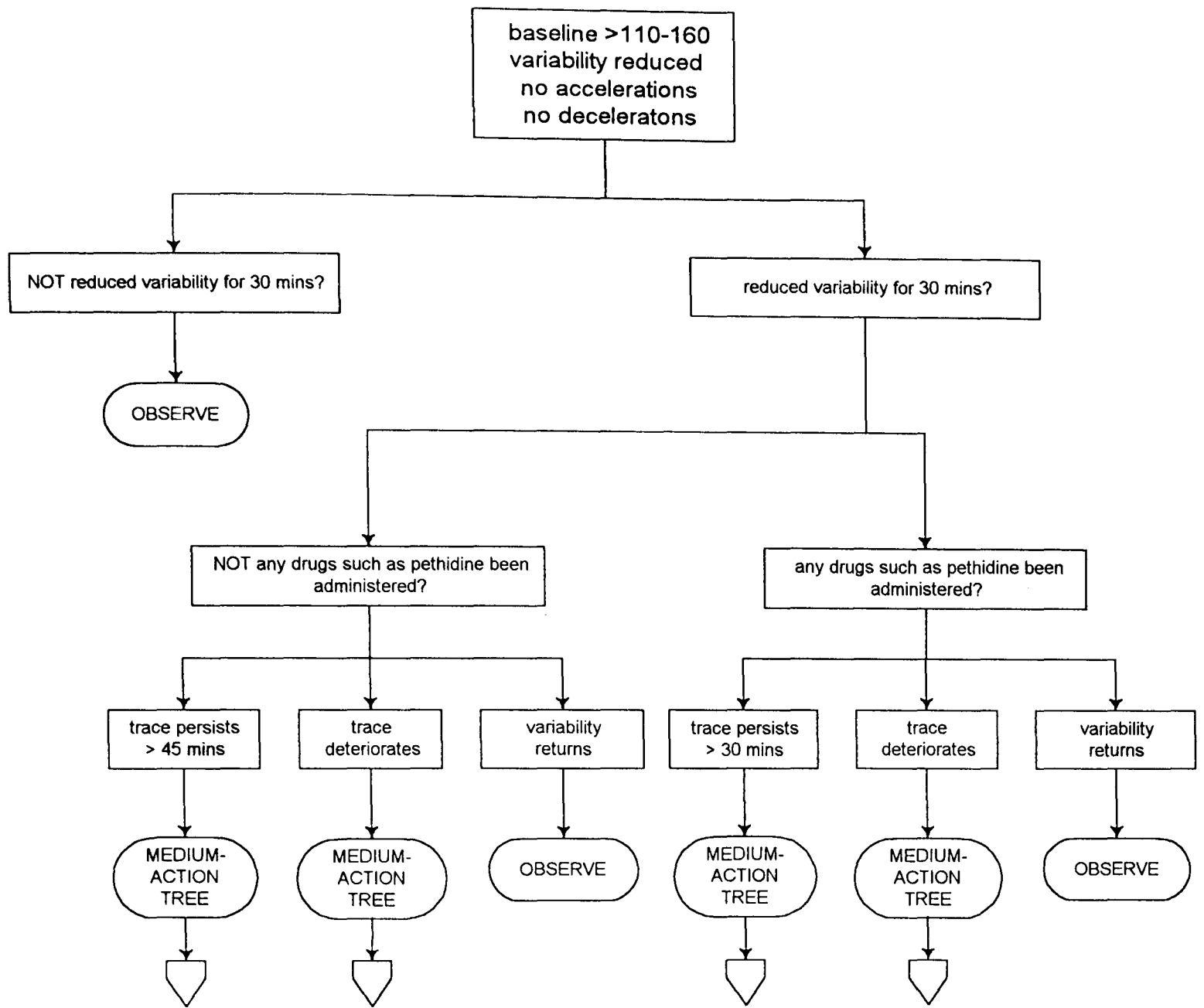
- variability goes from absent to at least reduced for 10 mins AND late/severe decelerations goto earlies or variables for at least 10 mins



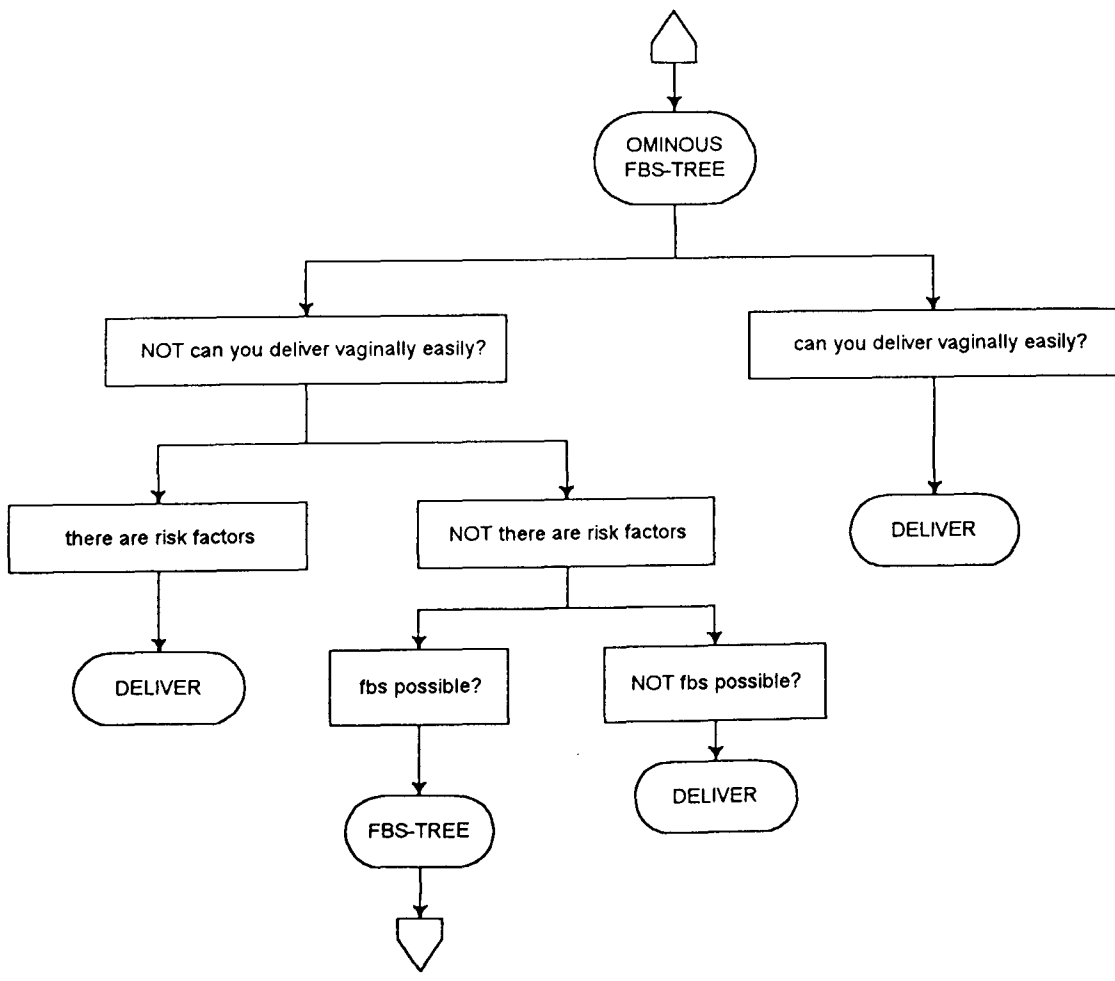
- trace becomes acceptable if baseline normal and variability at least reduced
- trace deteriorates if late or severe decelerations appear

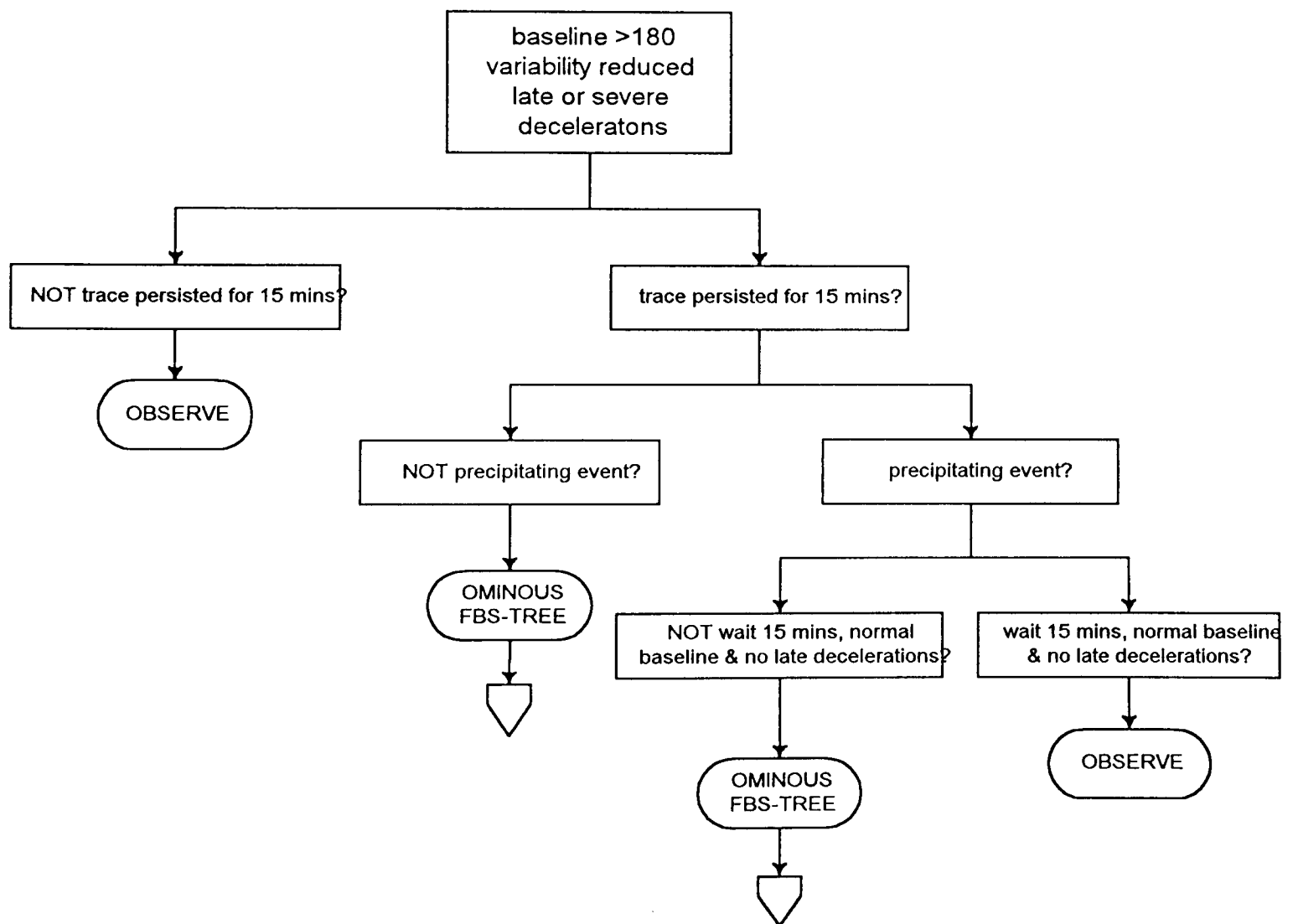
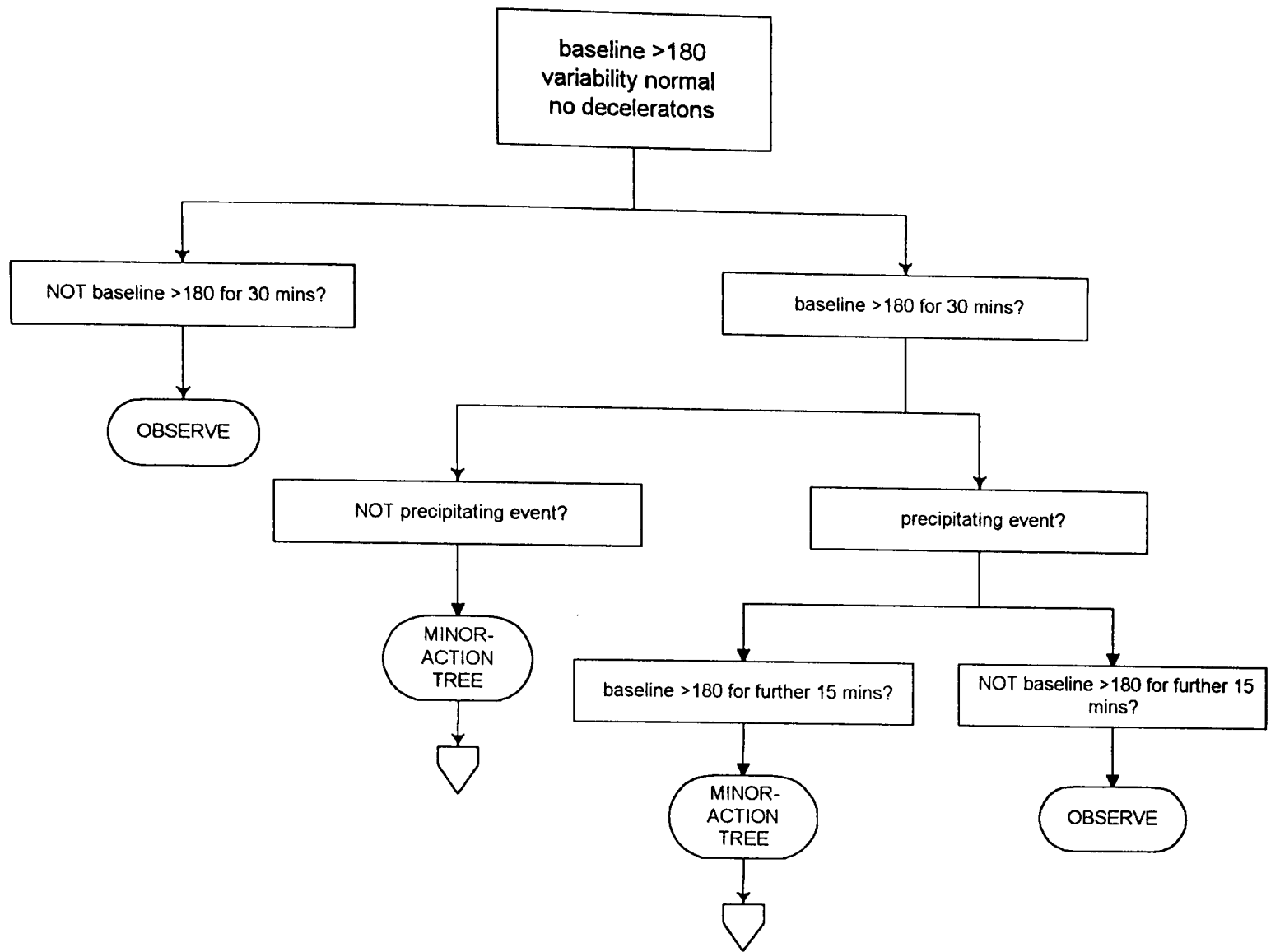


- trace becomes acceptable if no late decelerations for 10 mins
- trace deteriorates if baseline changes of variability becomes other than normal

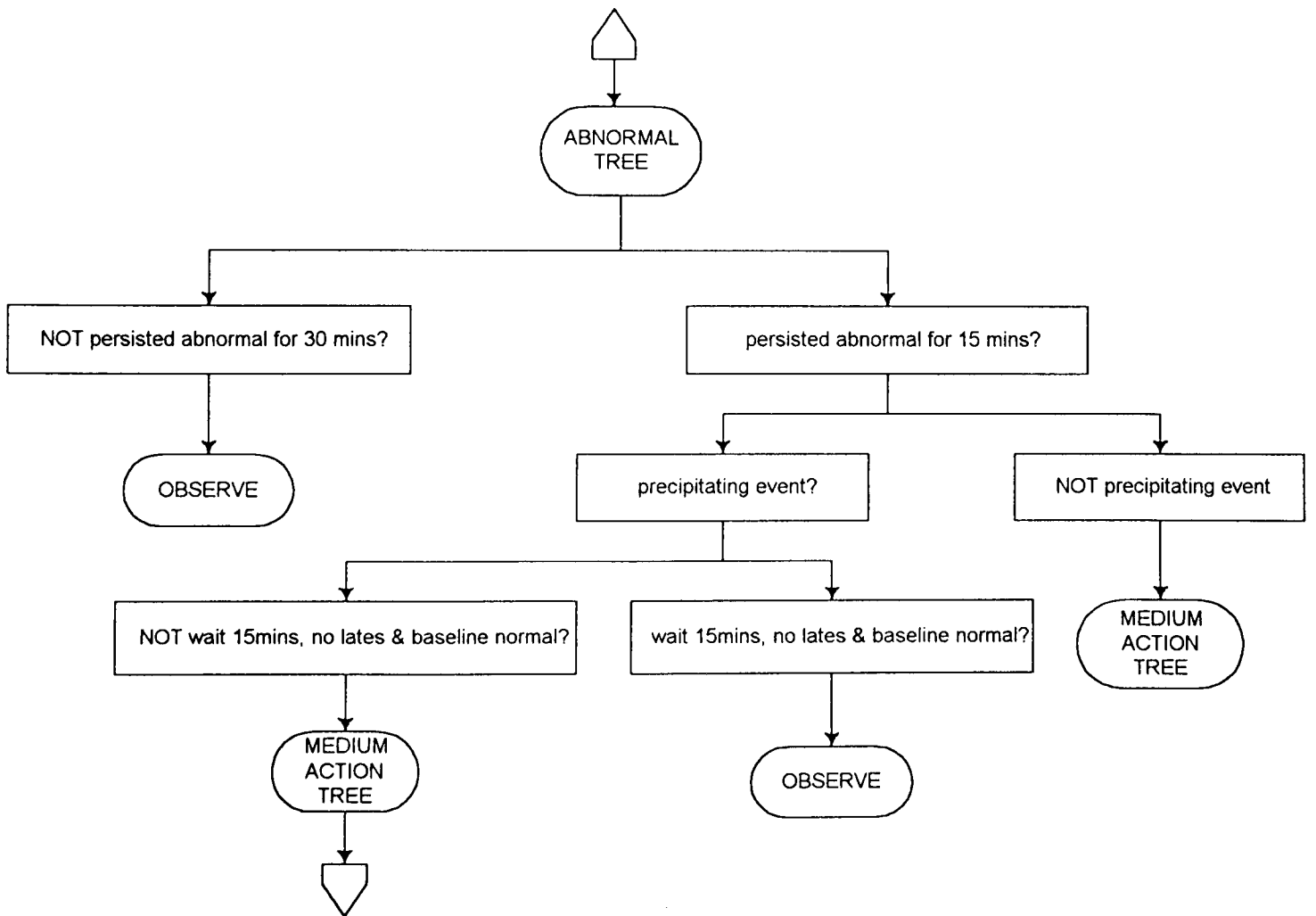
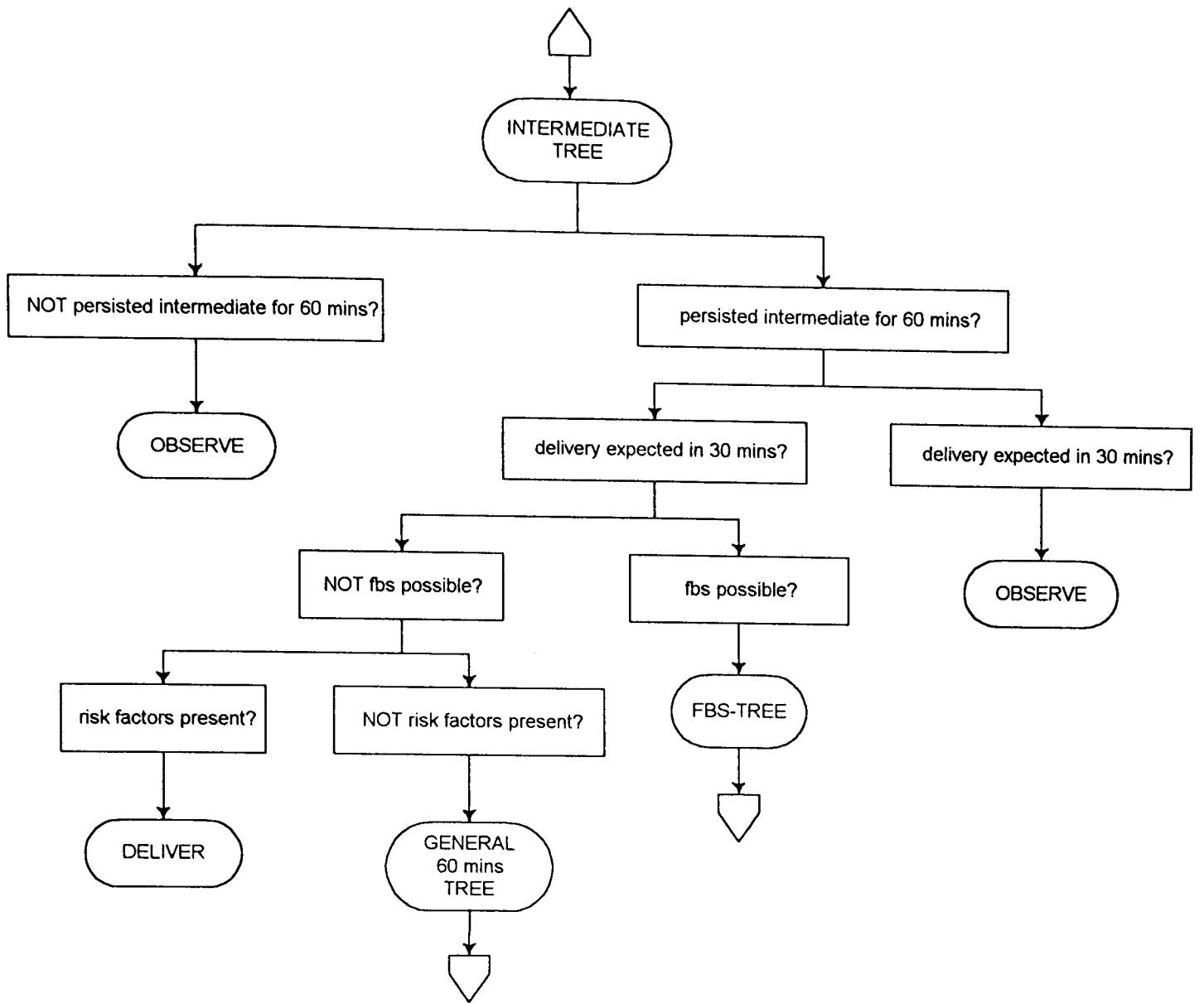


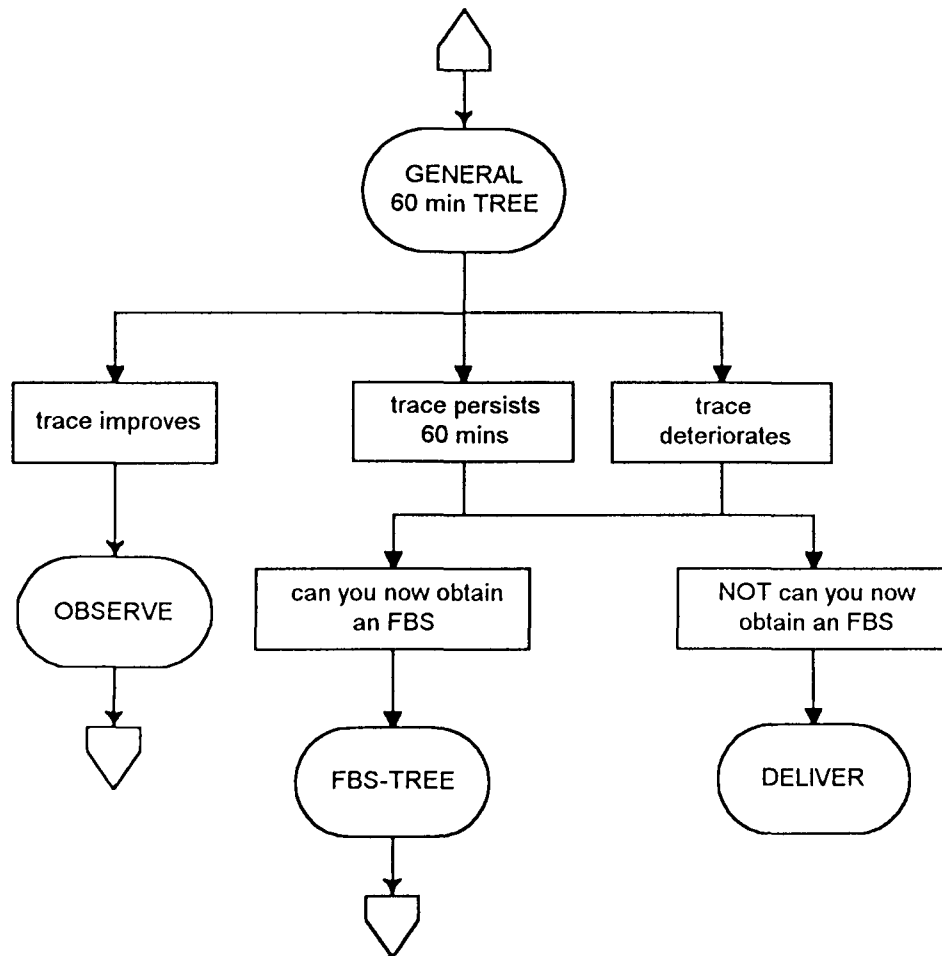
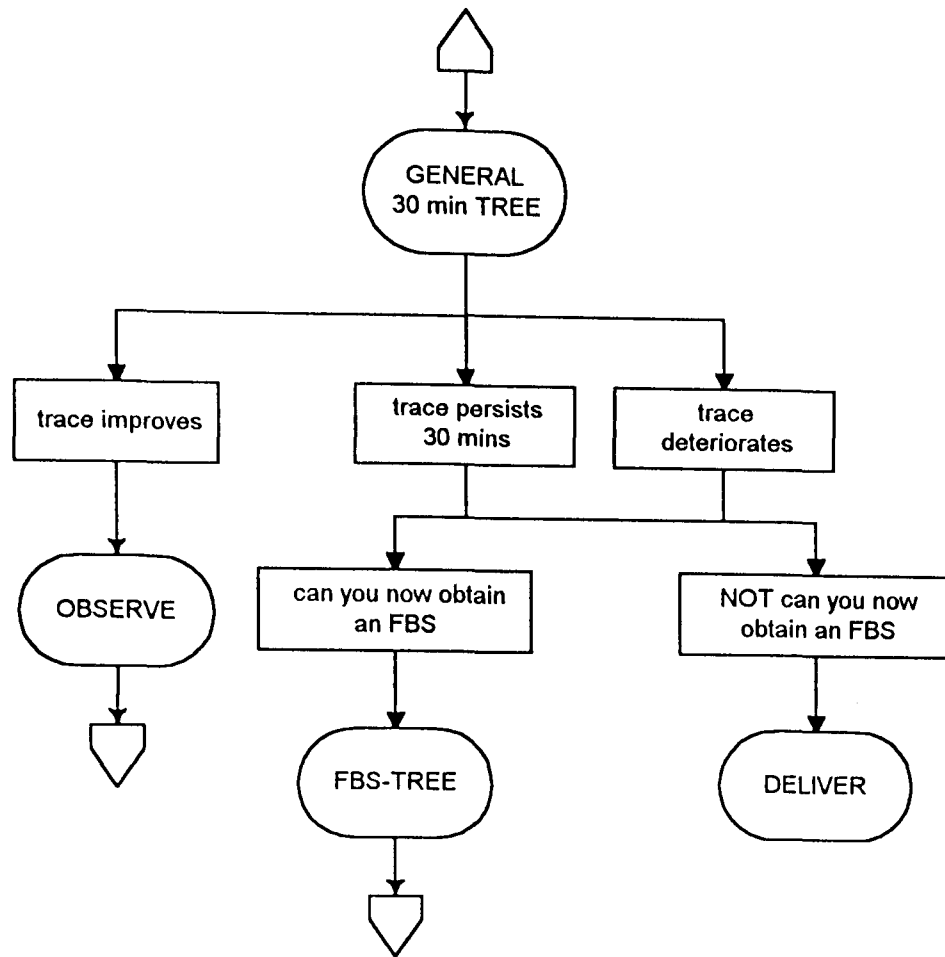
- trace persists if NOT variability returns AND NOT trace deteriorates
- trace deteriorates if late or severe decelerations appear or baseline changes



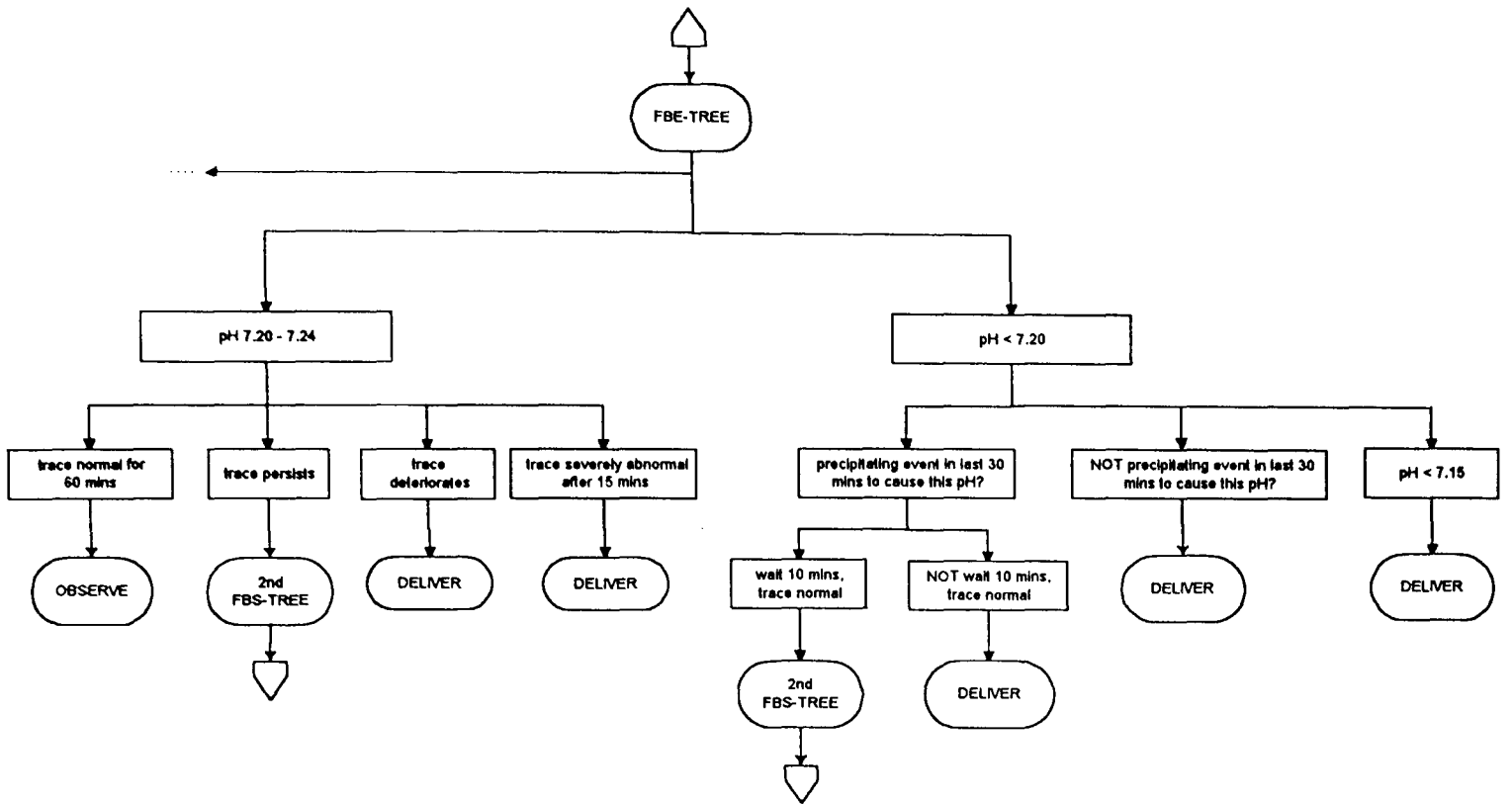


- has the trace persisted severely abnormal for 15 mins



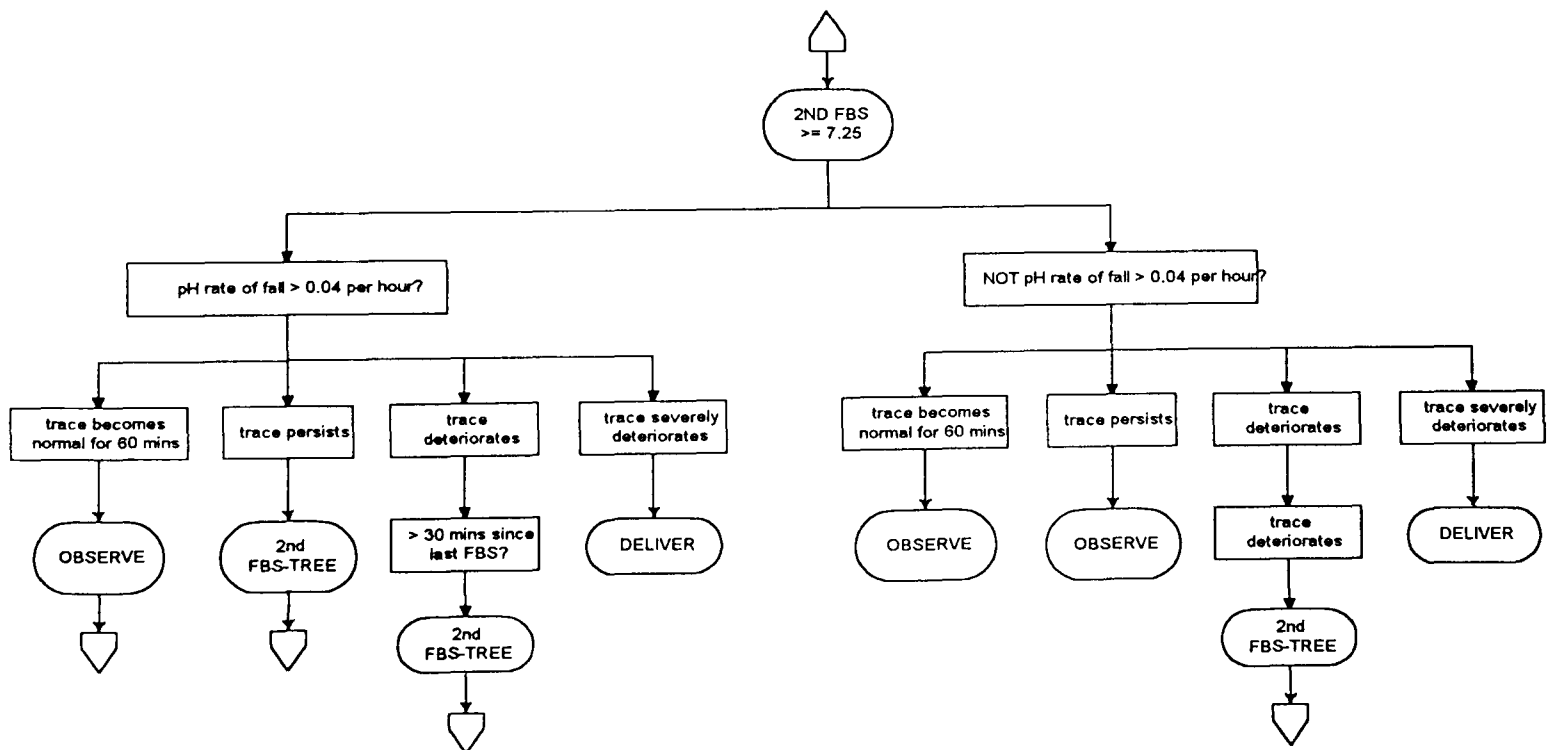


- trace deteriorates if it becomes severely abnormal
- trace improves if it becomes normal or intermediate
- trace persists if trace has not improved or trace has not deteriorated

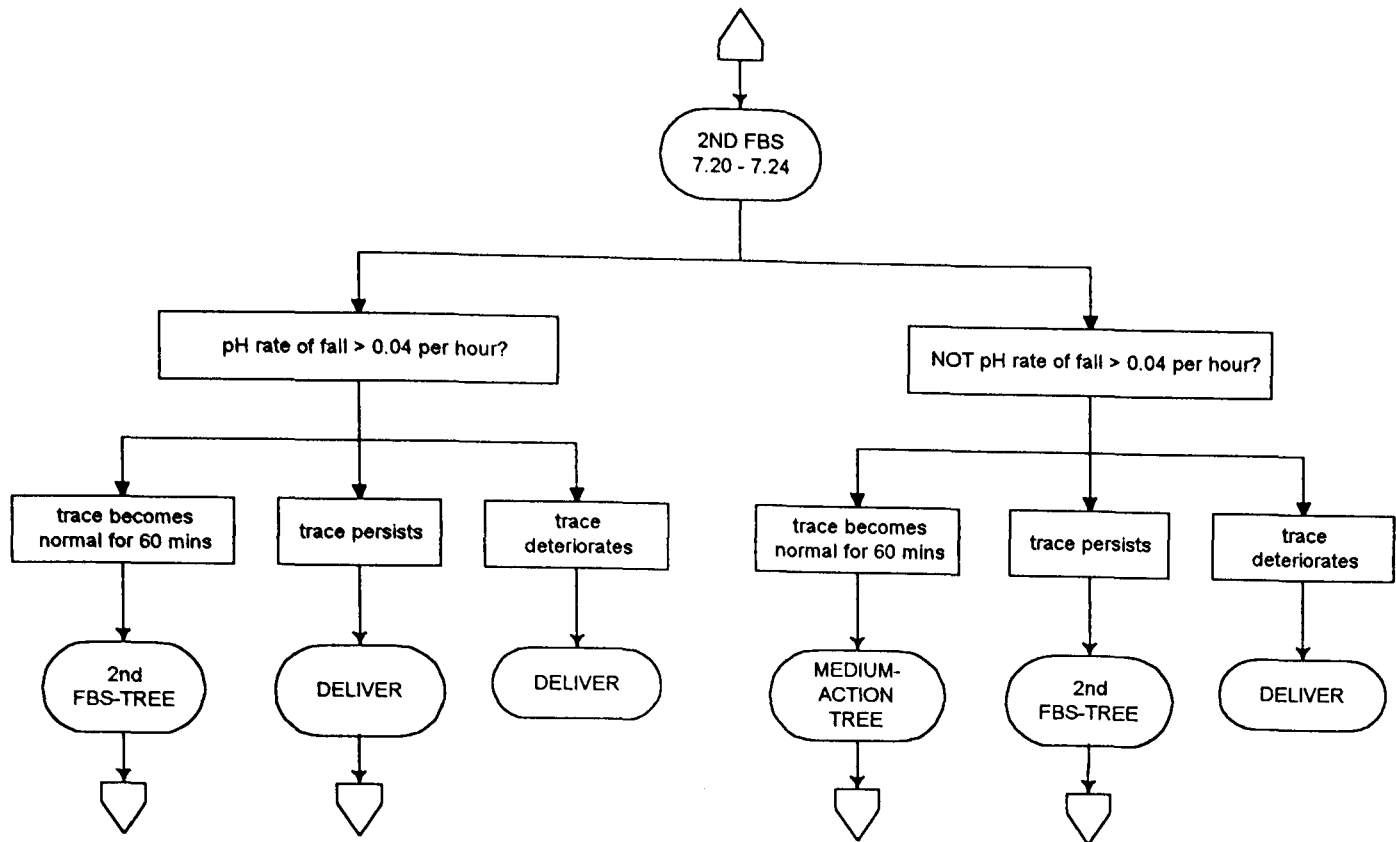


- trace deteriorates if it drops one classification
- trace normal for 60 mins if trace does not deteriorate in first 30 mins and is completely normal for second 30 mins
- trace persists if not normal for 60 mins and trace not deteriorated and trace not severely abnormal after 15 mins

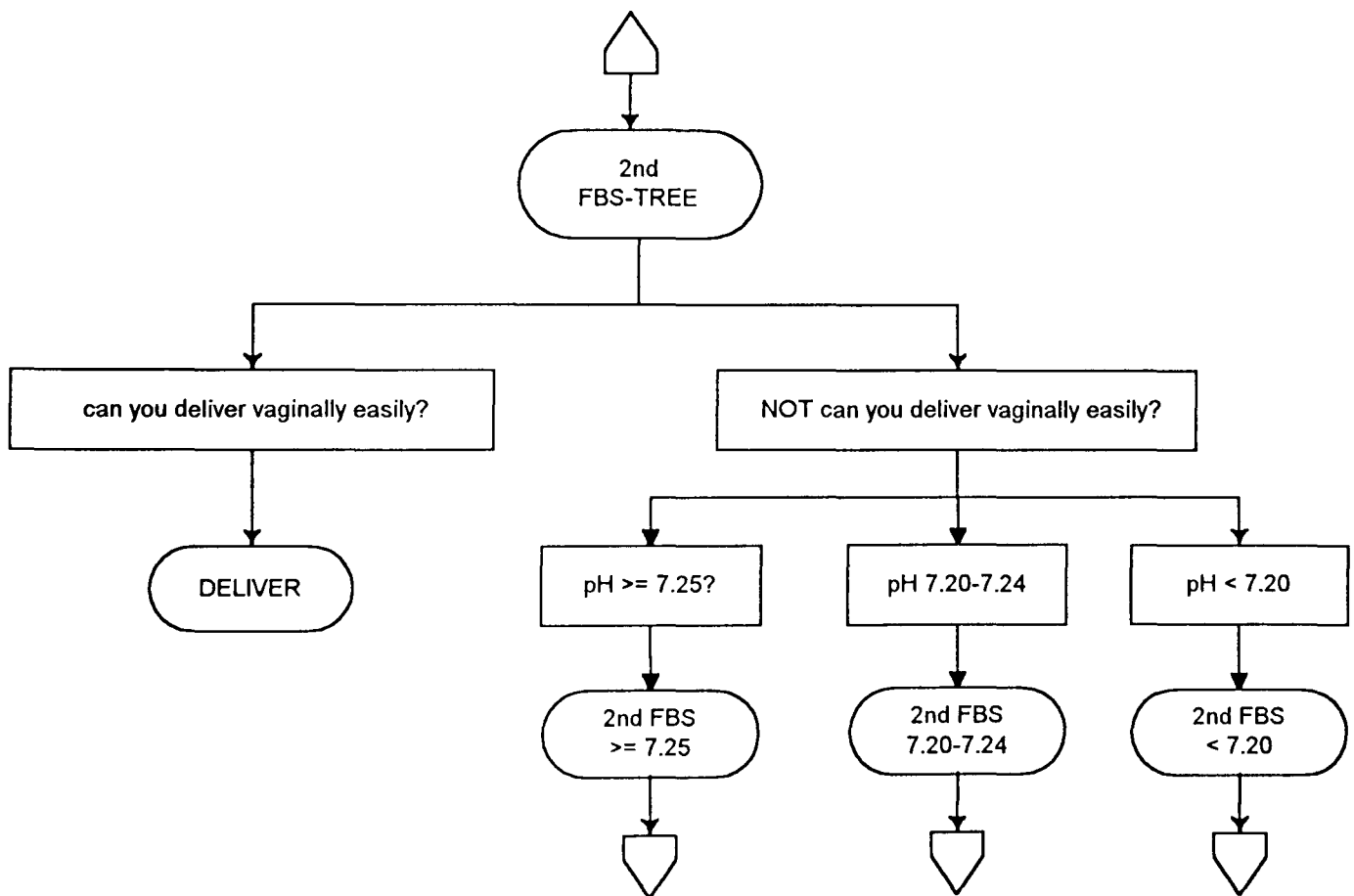
- a precipitating event in the last 30 mins which could have caused the low pH would be a severe bradycardia or severe decelerations in the last 15 mins without any previously

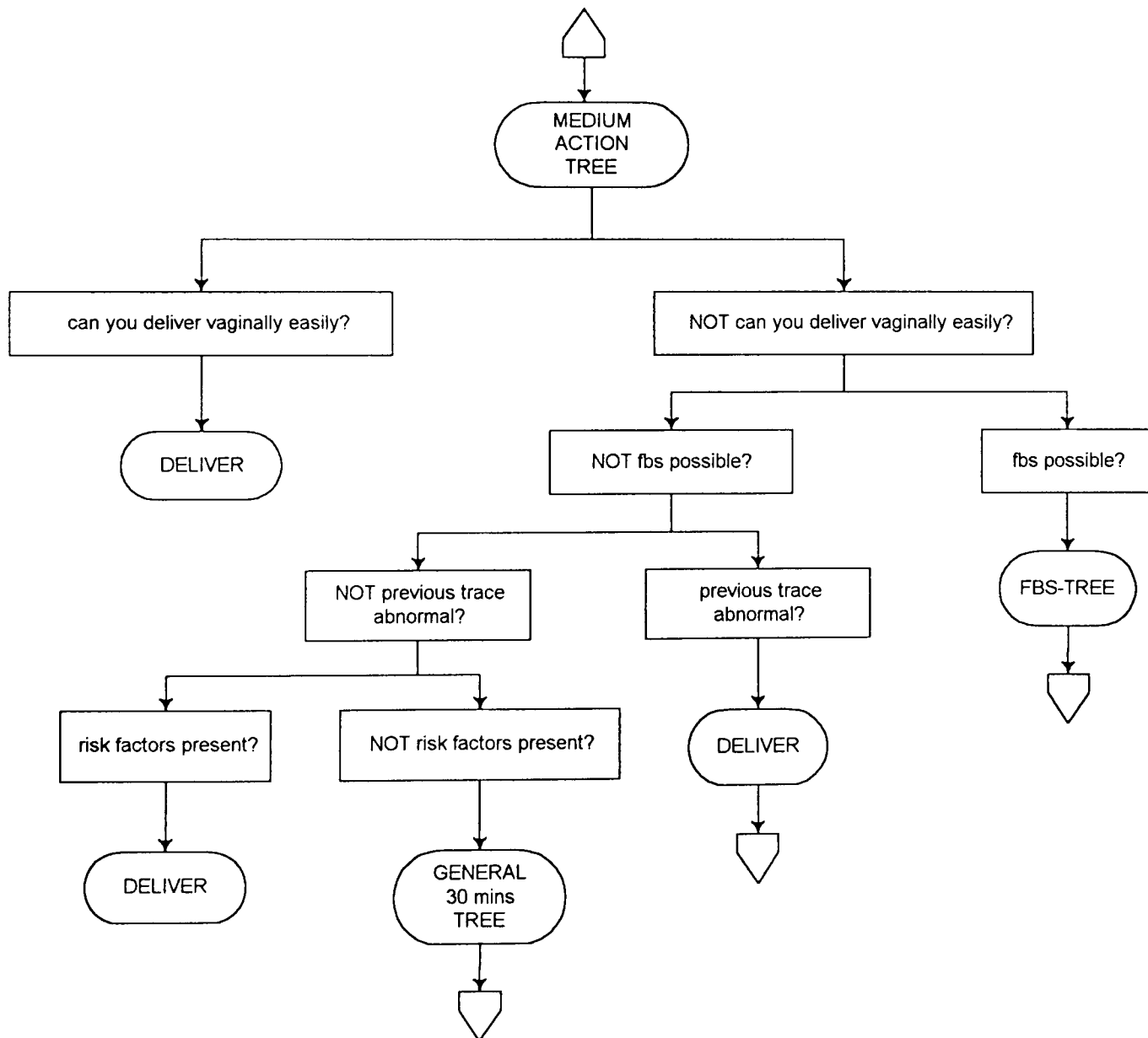
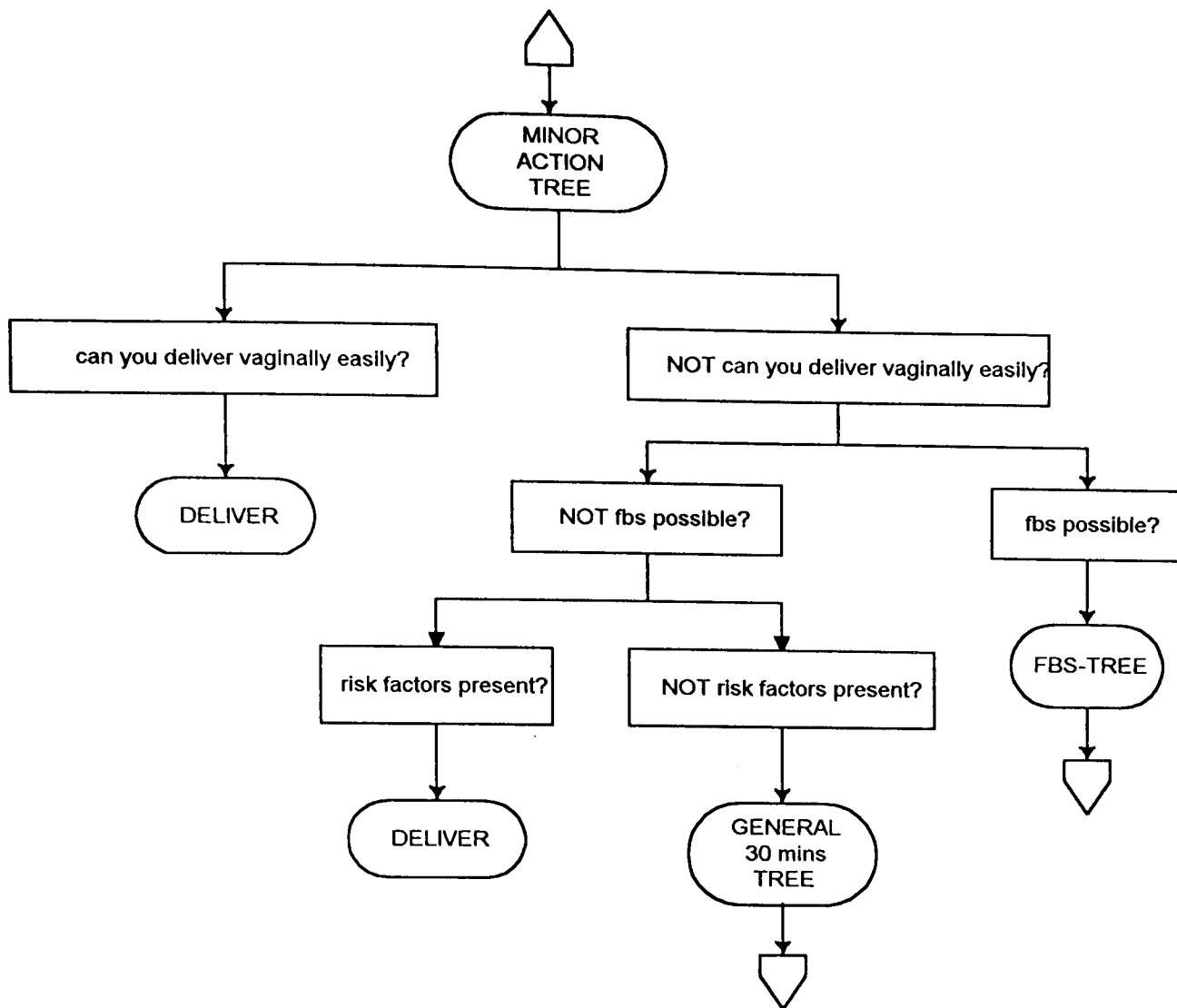


- trace deteriorates if it drops one classification
- trace becomes normal if trace returns to a classification which never requires and FBS for 30 mins in last 60 mins
- trace severely deteriorates if baseline < 90 or baseline now 90-110 and was > 180
- trace persists if trace does not deteriorate and trace does not become normal



- trace deteriorates if it drops one classification
- trace becomes normal for 60 mins if trace does not deteriorate and last 30 mins are all normal
- trace persists if trace does not deteriorate and trace does not become normal





Appendix E

The knowledge base.

Rule format:

rule(rule number, current node, next node, conditions)

rule(1,labour,<90,1)
rule(2,<90,prev_n,41)
rule(3,<90,prev_not_n,42)
rule(4,prev_n,base_imp,43)
rule(5,prev_n,goal_5,44)
rule(6,base_imp,goal_6,45)
rule(7,base_imp,goal_7,46,49)
rule(8,base_imp,fbt-tree,46,50,300)
rule(9,base_imp,goal_9,46,50,301)
rule(10,prev_not_n,pnn_base_imp,43)
rule(11,prev_not_n,goal_10,44)
rule(12,pnn_base_imp,goal_11,45)
rule(13,pnn_base_imp,goal_12,46,49)
rule(14,pnn_base_imp,fbt-tree,46,50,300)
rule(15,pnn_base_imp,goal_14,46,50,301)
rule(30,labour,goal_30,2,13,31,21)
rule(31,labour,goal_31,4,13,31,21)
rule(32,labour,goal_32,3,12,31,21)
rule(33,labour,goal_33,3,13,31,22)
rule(34,labour,goal_34,3,13,32,22)
rule(35,labour,goal_35,3,13,31,21)
rule(36,labour,goal_36,3,13,32,21)
rule(37,labour,goal_37,4,13,31,22)
rule(38,labour,goal_38,2,13,31,22)
rule(45,labour,intermediate,2,13,21,32)
rule(46,labour,intermediate,2,13,22,32)
rule(47,labour,intermediate,2,12,21,31)
rule(48,labour,intermediate,2,14,21,31)
rule(49,labour,intermediate,2,12,22,31)
rule(50,labour,intermediate,2,14,22,31)
rule(51,labour,intermediate,2,12,22,32)
rule(52,labour,intermediate,2,14,22,32)
rule(53,labour,intermediate,4,13,21,32)
rule(54,labour,intermediate,4,13,22,32)
rule(55,labour,intermediate,4,12,31)
rule(56,labour,intermediate,4,14,31)
rule(57,labour,intermediate,4,12,32)
rule(58,labour,intermediate,4,14,32)
rule(59,labour,intermediate,3,12,22,32)
rule(60,labour,intermediate,3,14,22,32)
rule(61,labour,intermediate,3,12,21,32)
rule(62,labour,intermediate,3,14,21,32)
rule(63,labour,intermediate,2,12,21,32)
rule(64,labour,intermediate,3,13,31)
rule(70,labour,abnormal,2,13,21,33)
rule(71,labour,abnormal,2,13,22,33)
rule(72,labour,abnormal,3,12,21,33)
rule(73,labour,abnormal,3,14,21,33)
rule(74,labour,abnormal,4,13,21,33)

rule(80,labour,severe-ab,2,12,21,33)
 rule(81,labour,severe-ab,2,14,21,33)
 rule(82,labour,severe-ab,2,12,22,33)
 rule(83,labour,severe-ab,2,14,22,33)
 rule(84,labour,severe-ab,4,13,22,33)
 rule(85,labour,severe-ab,4,12,33)
 rule(86,labour,severe-ab,4,14,33)
 rule(87,labour,severe-ab,3,12,22,33)
 rule(88,labour,severe-ab,3,14,22,33)
 rule(100,labour,var abs-no lates,2,11,31)
 rule(101,labour,var abs-no lates,2,11,32)
 rule(102,labour,var abs-no lates,4,11,31)
 rule(103,labour,var abs-no lates,4,11,32)
 rule(104,labour,var abs-no lates,3,11,32)
 rule(105,labour,var abs-no lates,3,11,31)
 rule(110,var abs-no lates,va-nl peth,51)
 rule(111,var abs-no lates,va-nl nopeth,52)
 rule(112,va-nl peth,va-nl peth menu,1000)
 rule(113,va-nl peth menu,goal_113,1002)
 rule(114,va-nl peth menu,medium-action,1004)
 rule(115,va-nl peth menu,medium-action,1003)
 rule(120,va-nl nopeth,va-nl nopeth menu,1001)
 rule(121,va-nl nopeth menu,goal_121,1002)
 rule(122,va-nl nopeth menu,medium-action,1005)
 rule(123,va-nl nopeth menu,medium-action,1003)
 rule(130,labour,var red-no decs,3,12,22,31)
 rule(135,var red-no decs,vr-nd ->30,65)
 rule(136,var red-no decs,goal_136,66)
 rule(137,vr-nd ->30,vr-nd ->30 peth,51)
 rule(138,vr-nd ->30,vr-nd ->30 nopeth,52)
 rule(139,vr-nd ->30 peth,vr-nd p menu,1010)
 rule(140,vr-nd ->30 nopeth,vr-nd no-p menu,1011)
 rule(141,vr-nd no-p menu,goal_141,1002)
 rule(142,vr-nd no-p menu,medium-action,1014)
 rule(143,vr-nd no-p menu,medium-action,1012)
 rule(144,vr-nd p menu,goal_144,1002)
 rule(145,vr-nd p menu,medium-action,1013)
 rule(146,vr-nd p menu,medium-action,1012)
 rule(150,labour,>180 all-n,5,13,31)
 rule(155,>180 all-n,>180 all-n ->30,61)
 rule(156,>180 all-n,goal_156,62)
 rule(157,>180 all-n ->30,>180 all-n precip,53)
 rule(158,>180 all-n ->30,minor-action,54)
 rule(159,>180 all-n precip,minor-action,63)
 rule(160,>180 all-n precip,goal_160,64)
 rule(170,labour,>180+,5,13,32)
 rule(171,labour,>180+,5,13,33)
 rule(172,labour,>180+,5,12,31)
 rule(173,labour,>180+,5,14,31)
 rule(174,labour,>180+,5,12,32)
 rule(175,labour,>180+,5,14,32)
 rule(180,>180+,>180+ ->15,515)

rule(181,>180+,goal_181,516)
 rule(182,>180+ ->15,>180+ precip,53)
 rule(183,>180+ ->15,medium-action,54)
 rule(184,>180+ precip,goal_184,55)
 rule(185,>180+ precip,medium-action,56)
 rule(190,labour,>180 vr-dl,5,12,33)
 rule(192,>180 vr-dl,>180 vr-dl ->15,515)
 rule(193,>180 vr-dl,goal_193,516)
 rule(194,>180 vr-dl ->15,>180 vr-dl precip,53)
 rule(195,>180 vr-dl ->15,ominous-fbs,54)
 rule(196,>180 vr-dl precip,goal_196,57)
 rule(197,>180 vr-dl precip,ominous-fbs,58)
 rule(200,labour,va-dl,2,11,33)
 rule(201,labour,va-dl,4,11,33)
 rule(202,labour,va-dl,5,11,33)
 rule(203,labour,va-dl,3,11,33)
 rule(205,va-dl,goal_205,59)
 rule(206,va-dl,ominous-fbs,60)
 rule(210,labour,>180 va-nl,5,11,31)
 rule(211,labour,>180 va-nl,5,11,32)
 rule(213,>180 va-nl,va-nl peth,51)
 rule(214,>180 va-nl,>180 va-nl menu,1020)
 rule(215,>180 va-nl menu,goal_215,1021)
 rule(216,>180 va-nl menu,medium-action,1022)
 rule(217,>180 va-nl menu,medium-action,1023)
 rule(220,labour,all n-ld,3,13,21,33)
 rule(221,labour,all n-ld,3,13,22,33)
 rule(222,all n-ld,all n-ld ->30,530)
 rule(223,all n-ld,goal_221,531)
 rule(224,all n-ld ->30,goal_222,47)
 rule(225,all n-ld ->30,all n-ld pre,48,53)
 rule(226,all n-ld ->30,all n-ld nopre,48,54)
 rule(227,all n-ld pre,goal_225,57)
 rule(228,all n-ld pre,all n-ld nopre,58)
 rule(229,all n-ld nopre,fbs-tree,300)
 rule(230,all n-ld nopre,all n-ld risk,301,200)
 rule(231,all n-ld nopre,all n-ld nrisk,301,201)
 rule(232,all n-ld nrisk,general 60 menu,1600)
 rule(234,all n-ld risk,general 30 menu,1500)
 rule(240,labour,vr-ld,3,12,22,33)
 rule(245,vr-ld,vr-ld assess dilat,400)
 rule(246,vr-ld assess dilat,abnormal,401)
 rule(247,vr-ld assess dilat,vr-ld 8-10cm,402)
 rule(248,vr-ld assess dilat,vr-ld 2nd stage,403)
 rule(249,vr-ld 8-10cm,vr-ld 8-10cm ->30,530)
 rule(250,vr-ld 8-10cm,goal_250,531)
 rule(251,vr-ld 8-10cm ->30,goal_251,47)
 rule(252,vr-ld 8-10cm ->30,8-10cm precip,48,53)
 rule(253,vr-ld 8-10cm ->30,8-10cm nopre,48,54)
 rule(254,8-10cm precip,goal_254,57)
 rule(255,8-10cm precip,8-10cm nopre,58)
 rule(256,8-10cm nopre,fbs-tree,300)

rule(257,8-10cm nopre,general 30 menu,301,1500)
 rule(260,vr-ld 2nd stage,vr-ld 2nd ->30,530)
 rule(261,vr-ld 2nd stage,goal_261,531)
 rule(262,vr-ld 2nd ->30,goal_262,47)
 rule(263,vr-ld 2nd ->30,goal_263,48,80)
 rule(264,vr-ld 2nd ->30,vr-ld no-forc,48,81)
 rule(265,vr-ld no-for,fbs-tree,300)
 rule(266,vr-ld no-for,general 30 menu,301,1500)
 rule(300,minor-action,goal_300,49)
 rule(301,minor-action,fbs-tree,50,300)
 rule(302,minor-action,goal_302,50,301,200)
 rule(303,minor-action,min-act no-risk,50,301,201)
 rule(304,min-act no-risk,general 30 menu,1500)
 rule(320,medium-action,goal_320,49)
 rule(321,medium-action,fbs-tree,50,300)
 rule(322,medium-action,goal_322,50,301,67)
 rule(323,medium-action,goal_323,50,301,68,200)
 rule(324,medium-action,med-a no-r,50,301,68,201)
 rule(325,med-a no-r,general 30 menu,1500)
 rule(340,ominous-fbs,goal_340,49)
 rule(341,ominous-fbs,goal_341,50,200)
 rule(342,ominous-fbs,fbs-tree,50,201,300)
 rule(343,ominous-fbs,goal_343,50,201,301)
 rule(360,intermediate,goal_360,560,82)
 rule(361,intermediate,goal_361,561)
 rule(362,intermediate,fbs-tree,560,83,300)
 rule(363,intermediate,goal_363,560,83,301,200)
 rule(364,intermediate,int no-r,560,83,301,201)
 rule(365,int no-r,general 60 menu,1600)
 rule(380,abnormal,abnormal ->30,530)
 rule(381,abnormal,goal_381,531)
 rule(382,abnormal ->30,abnormal precip,53)
 rule(383,abnormal ->30,medium-action,54)
 rule(384,abnormal precip,goal_384,57)
 rule(385,abnormal precip,medium-action,58)
 rule(390,severe-ab,severe-ab ->15,515)
 rule(391,severe-ab,goal_391,516)
 rule(392,severe-ab ->15,severe-ab precip,53)
 rule(393,severe-ab ->15,ominous-fbs,54)
 rule(394,severe-ab precip,goal_394,57)
 rule(395,severe-ab precip,ominous-fbs,58)
 rule(400,fbs-tree,>=7.25,325)
 rule(401,>=7.25,sev-ab fbs-menu,69)
 rule(402,>=7.25,fbs-menu,70)
 rule(405,sev-ab fbs-menu,goal_405,1100,73)
 rule(406,sev-ab fbs-menu,2nd-fbs,1100,351)
 rule(407,sev-ab fbs-menu,goal_407,1100,77,87)
 rule(408,sev-ab fbs-menu,2nd-fbs,1100,77,88,352)
 rule(410,fbs-menu,goal_410,1101,73)
 rule(411,fbs-menu,2nd-fbs,1101,350)
 rule(412,fbs-menu,fbs-menu,1101,75,71,370)
 rule(413,fbs-menu,2nd-fbs,1101,75,72)

rule(414, fbs-menu, goal_414, 1101, 77, 87)
 rule(415, fbs-menu, 2nd-fbs, 1101, 88, 352)
 rule(420, fbs-tree, 7.20-7.24_1, 326)
 rule(421, 7.20-7.24_1, goal_421, 1110, 84)
 rule(422, 7.20-7.24_1, 2nd-fbs, 1110, 85)
 rule(423, 7.20-7.24_1, goal_423, 1110, 75)
 rule(424, 7.20-7.24_1, goal_424, 1110, 86)
 rule(430, fbs-tree, < 7.20, 327)
 rule(431, < 7.20, goal_431, 327, 328)
 rule(432, < 7.20, 2nd-fbs, 91, 93)
 rule(433, < 7.20, goal_433, 91, 94)
 rule(434, < 7.20, goal_434, 92)
 rule(450, 2nd-fbs, goal_450, 89)
 rule(451, 2nd-fbs, >=7.25_2, 90, 330)
 rule(452, 2nd-fbs, 7.20-7.24_2, 90, 331)
 rule(453, 2nd-fbs, goal_453, 90, 332)
 rule(460, >=7.25_2, goal_460, 1130, 100, 102)
 rule(461, >=7.25_2, 2nd-fbs, 1130, 100, 103, 106)
 rule(462, >=7.25_2, 2nd-fbs, 1130, 100, 104, 110, 106)
 rule(463, >=7.25_2, goal_463, 1130, 100, 105)
 rule(465, >=7.25_2, goal_465, 1130, 101, 102)
 rule(466, >=7.25_2, goal_466, 1130, 101, 103)
 rule(467, >=7.25_2, 2nd-fbs, 1130, 101, 104, 110, 106)
 rule(468, >=7.25_2, goal_468, 1130, 101, 105)
 rule(470, 7.20-7.24_2, 2nd-fbs, 1120, 100, 102, 106)
 rule(471, 7.20-7.24_2, goal_471, 1120, 100, 103)
 rule(472, 7.20-7.24_2, goal_472, 1120, 100, 104)
 rule(475, 7.20-7.24_2, goal_475, 1120, 101, 102)
 rule(476, 7.20-7.24_2, 2nd-fbs, 1120, 101, 103, 106)
 rule(477, 7.20-7.24_2, goal_477, 1120, 101, 104)
 rule(500, general 30 menu, goal_500, 1032)
 rule(501, general 30 menu, try fbs again, 1035)
 rule(502, general 30 menu, try fbs again, 1034)
 rule(510, general 60 menu, goal_510, 1032)
 rule(511, general 60 menu, try fbs again, 1033)
 rule(512, general 60 menu, try fbs again, 1034)
 rule(520, try fbs again, fbs-tree, 302)
 rule(521, try fbs again, goal_521, 303)
 rule(900, quality, labour, 900)
 rule(901, quality, goal_901, 901)
 cond(1, baseline < 90, 0)
 cond(2, baseline 90 - 110, 0)
 cond(3, baseline 110 - 160, 0)
 cond(4, baseline 160 - 180, 0)
 cond(5, baseline > 180, 0)
 cond(11, variability absent, 0)
 cond(12, variability reduced, 0)
 cond(13, variability normal, 0)
 cond(14, variability increased, 0)
 cond(21, acceleration present, 0)
 cond(22, not accerations present, 0)
 cond(31, no decs, 0)

cond(32,early or mild variable decs,0)
cond(33,late or severe variable decs,0)
cond(41,prev trace normal,0)
cond(42,not prev trace normal,0)
cond(43,action wait 10 baseline improv ?,43)
cond(44,not action wait 10 baseline improv ?,0)
cond(45,wait 10 baseline normal ?,6)
cond(46,not wait 10 baseline normal ?,0)
cond(47,do you expect delivery in next 15,47)
cond(48,not do you expect delivery in next 15,0)
cond(49,can deliver vaginally,49)
cond(50,not can deliver vaginally,0)
cond(51,pethidine question,51)
cond(52,not pethidine question,0)
cond(53,precipitating event question,53)
cond(54,not precipitating event question,0)
cond(55,wait 15 baseline normal ?,55)
cond(56,not wait 15 baseline normal ?,0)
cond(57,wait 15 norm base & no ld,57)
cond(58,not wait 15 norm base & no ld,0)
cond(59,wait 15 va to vr or better & no ld,59)
cond(60,not wait 15 va to vr or better & no ld,0)
cond(61,baseline >180 for 30 mins,61)
cond(62,not baseline >180 for 30 mins,62)
cond(63,baseline >180 for further 15 mins,63)
cond(64,not baseline >180 for further 15 mins,0)
cond(65,variability red >=30 mins,65)
cond(66,not variability red >=30 mins,0)
cond(67,the previous trace was abnormal,67)
cond(68,not the previous trace was abnormal,0)
cond(69,previous trace was sev-abnormal,69)
cond(70,not previous trace was sev-abnormal,0)
cond(71,wait 30, trace improves one category,71)
cond(72,not wait 30,trace improves one cat,0)
cond(73,trace is now normal,90)
cond(75,trace drops one category,75)
cond(77,trace severely deteriorated,77)
cond(80,can do easy forceps,80)
cond(81,not can do easy forceps,81)
cond(82,delivery expected in 30 mins,82)
cond(83,not delivery expected in 30 mins,0)
cond(84,trace normal for 60 mins,84)
cond(85,trace persists for 60 mins,85)
cond(86,trace sev-ab after 15 mins,86)
cond(87,can you now deliver vaginally_2,87)
cond(88,not can you now deliver vaginally_2,0)
cond(89,can you now deliver vaginally_3,89)
cond(90,not can you now deliver vaginally_3,0)
cond(91,there was a precip causing low pH,91)
cond(92,not there was a precip causing low pH,0)
cond(93,wait 10 min trace is normal,93)
cond(94,not wait 10 min trace is normal,0)

cond(100,rate of fall in pH >0.04/hr,100)
cond(101,not rate of fall in pH >0.04/hr,101)
cond(102,fbs2: trace normal 60 min,102)
cond(103,fbs2: trace persists 60',103)
cond(104,fbs2: trace deteriorates,104)
cond(105,fbs2: trace sev deteriorates,105)
cond(106,fbs2: clear conditions since 2nd-fbs,0)
cond(110,fbs2: 30 mins since last fbs,110)
cond(200,there are risk factors,200)
cond(201,not there are risk factors,0)
cond(300,can you do an fbs,300)
cond(301,not can you do an fbs,0)
cond(302,can you now do an fbs,302)
cond(303,not can you now do an fbs,303)
cond(325,fbs_1 result >= 7.25,0)
cond(326,fbs_1 result 7.20 - 7.24,0)
cond(327,fbs_1 result < 7.20,0)
cond(328,fbs_1 result < 7.15,0)
cond(330,fbs_2 result >= 7.25,0)
cond(331,fbs_2 result 7.20 - 7.24,0)
cond(332,fbs_2 result < 7.20,0)
cond(333,fbs_2 result < 7.15,0)
cond(350,60 mins since last fbs,350)
cond(351,30 mins since last fbs,351)
cond(352,15 mins since last fbs,352)
cond(370,reset fbs menu,0)
cond(400,cervical dilatation menu,400)
cond(401,cervical dilataion < 8 cm,0)
cond(402,cervical dilatation 8 - 10 cm,0)
cond(403,second stage,0)
cond(515,trace persists sev-ab >= 15 mins,515)
cond(516,not trace persists sev-ab >= 15 mins,0)
cond(530,trace persists ab >= 30 mins,530)
cond(531,not trace persists ab >= 30 mins,0)
cond(560,trace persists int >= 60 mins,560)
cond(561,not trace persists int >= 60 mins,0)
cond(900,the scalp quality is good,900)
cond(901,not the scalp quality is good,900)
cond(1000,abs var no lates peth trigger,1000)
cond(1001,abs var no lates nopeth trigger,1001)
cond(1002,variability now normal,0)
cond(1003,this trace has deteriorated,0)
cond(1004,this trace persisted >= 60 mins,0)
cond(1005,this trace persisted >= 30 mins,0)
cond(1010,var red no decs trigger peth,1010)
cond(1011,var red no decs trigger no peth,1011)
cond(1012,this trace has deteriorated,0)
cond(1013,this tace persisted >=45 mins,0)
cond(1014,this trace persisted >=30 mins,0)
cond(1020,>180 va-nl trigger,1020)
cond(1021,trace becomes acceptable,0)
cond(1022,trace persists >=15,0)

cond(1023,trace deteriorates,0)
cond(1032,trace now acceptable,0)
cond(1033,trace persists >= 60 mins,0)
cond(1034,trace deteriorates,0)
cond(1035,trace persists >= 30 mins,0)
cond(1100,sev-ab fbs menu trigger,1100)
cond(1101,fbs menu trigger,1101)
cond(1110,fbs 7.2 - 7.24 trigger,1110)
cond(1120, 2nd fbs 7.2 - 7.24 trigger,1120)
cond(1130, 2nd fbs 7.2 - 7.24 trigger,1130)
cond(1500,general 30 trigger,1500)
cond(1600,general 60 trigger,1600)

Appendix F

An expert system written in Prolog software.

/* An Expert system written in Prolog for the management of labour */

/*- R.D.F Keith Feb 1991 */

```
code = 2000

global predicates
inf() - language c
mainy - language c

global domains
file = myfile
CONDITIONS = BNO*
HISTORY = RNO*
RNO, BNO, FNO = INTEGER
CATEGORY = STRING
RISK = string
data_file = string
slist = string*

GLOBAL DATABASE
rule(RNO, CATEGORY, CATEGORY, CONDITIONS)
cond(BNO, STRING)
data_file(data_file)
yes(BNO)
no(BNO)
fact(FNO, CATEGORY, CATEGORY)
topic(string)
risk_factor(string)

include "tdoms.pro"
include "tpreds.pro"
include "menu2.pro"

PREDICATES
/*Commands*/
title_go
load_know
erase
clear
proces(integer)
endd(integer)
goes(CATEGORY)
run
reverse(CONDITIONS, CONDITIONS)
reverse1(CONDITIONS, CONDITIONS, CONDITIONS)
write_data(integer)

/*patient_info*/

/*Inferences mechanisms*/
go(HISTORY, CATEGORY)

check(RNO, HISTORY, CONDITIONS)
notest(BNO)
inpq(HISTORY, RNO, BNO, STRING)
do_answer(HISTORY, RNO, STRING, BNO, INTEGER)
norm_ctg
```

```

assert_risk_cond(string)
add_risk

labour_risk_info
antenatal_risk_info
choice(integer)
stop(integer)
choi(integer)
sto(integer)
trace_assess
baseline_assess
proces1(integer)
stop1(integer)
variability_assess
proces2(integer)
stop2(integer)
accelerations_assess
proces3(integer)
stop3(integer)
decelerations_assess
proces4(integer)
stop4(integer)
fbs_menu
proces5(integer)
stop5(integer)
trace_condition_menu
proces6(integer)
stop6(integer)
fbs_menu_2
proces7(integer)
stop7(integer)
trace_condition_menu_2
proces8(integer)
stop8(integer)
trace_condition_menu_3
proces9(integer)
stop9(integer)
v_absent_d_nolates_nopeth_menu
proces10(integer)
stop10(integer)
v_absent_d_nolates_peth_menu
proces11(integer)
stop11(integer)
v_absent_d_lates_menu
proces12(integer)
stop12(integer)
cervical_dilatation_menu
proces13(integer)
stop13(integer)

```

GOAL

```

makewindow(1,14,25,"",4,0,20,80), /* inner working window */
makewindow(2,0,31,"",14,0,10,80),
makewindow(5,14,0,"",0,0,3,80),
makewindow(7,15,0,"",3,17,1,45),
makewindow(8,79,0,"",24,0,1,80), /* boarder with action text in */
makewindow(9,12,0,"",0,0,25,80),
    openwrite(myfile,"labour.dat"),

    /*patient_info,*/
    TIME = 0,

```



```

run.
clauses
run :-
  shiftwindow(7),
  clearwindow,
  write("    INtelligent Fetal AssessmeNT (INFANT)"),
  shiftwindow(8),
  clearwindow,
  write("  select option with arrow key  "),
  antenatal_risk_info,
  load_know,
  repeat,
    shiftwindow(1),
    menu(6,55,15,9,
        ["Normal CTG",
         "Consultation",
         "Update info"],
        "menu",2,CHOICE),
    proces(CHOICE),
  endd(CHOICE),!.

proces(0):-shiftwindow(9), clearwindow, closefile(myfile), exit.
proces(1):-norm_ctg.
proces(2):-writedevicemyfile(myfile), write("0\n"), write("2\n"),

        writedevicemyfile(screen),
        title_go.
proces(3):-labour_risk_info.

endd(0):-shiftwindow(9), clearwindow, closefile(myfile).

/*patient_info:-write("Are there antenatal risk factors"),*/
/*      readln(_).*/

/*Inference mechanism*/
trace_assess:-
  baseline_assess,
  variability_assess,
  accelerations_assess,
  decelerations_assess.

title_go:-
  goes(Mygoal),
  nl,nl,go([],Mygoal),!.

goes(Mygoal):-
  clearwindow,
  clear,
  trace_assess,
  risk_factor(X),
  assert_risk_cond(X),

  Mygoal = "heart_rate".

assert_risk_cond(none):-
  assert(no(26)).

assert_risk_cond(present):-
  assert(yes(26)).

go( _, Mygoal ):-
  /* My best guess */

```

```

not(rule(_,Mygoal,_,_)),!,nl,
write("Recommended Action:\n\n",Mygoal),nl,nl,
writedevicemyfile), write("0\n"), write("Recommended action:- ",
Mygoal),
write("\n0\n"),
writedevicemyscreen),
write("\n\nPress any key to continue "),
readchar(_),
clearwindow.

go( HISTORY, Mygoal ):-
rule(RNO,Mygoal,NY,COND),
check(RNO,HISTORY, COND),
go([RNO|HISTORY],NY).

/** VARIABILITY ABSENT DECELERATIONS-NO LATES NO PETHIDINE **/
check( _, _, [74|_] ):-
assert(yes(74)),
clearwindow,
write("\nContinue to observe the trace until one of the following
become true"),
v_absent_d_nolates_nopeth_menu,
!.
/*****/

/** VARIABILITY ABSENT DECELERATIONS-NO LATES PETHIDINE **/
check( _, _, [79|_] ):-
assert(yes(79)),
clearwindow,
write("\nContinue to observe the trace until one of the following
become true"),
v_absent_d_nolates_peth_menu,
!.
/*****/

/** VARIABILITY ABSENT DECELERATIONS LATES ***/
check( _, _, [80|_] ):-
assert(yes(80)),
clearwindow,
write("\nContinue to observe the trace until one of the following
become true"),
v_absent_d_lates_menu,
!.
/*****/

/** ASSESS CERVICAL DILATATION ***/
check( _, _, [81|_] ):-
assert(yes(81)),
clearwindow,
write("\n\tWhat is the Cervical Dilatation"),
cervical_dilatation_menu,
!.
/*****/

/** FBS RESULT ***/
check( _, _, [100|_] ):-
assert(yes(100)),
clearwindow,
write("\n\tPerform an FBS and enter result"),
fbs_menu,
!.

```

```

/*****/

/***** Wait for 30 mins *****/
/** used in fbs tree to wait for 30 minutes before *****/
/* requesting another fbs when the trace is deteriorating */
check( _, _, [109|_] ):-

    retract(yes(106)),
    retract(no(105)),
    retract(no(107)),
    retract(yes(109)),
    fail.

/*****/

/***** ASSESS TRACE *****/
check( _, _, [104|_] ):-
    assert(yes(104)),
    clearwindow,
    write("\nObserve the trace:"),
    trace_condition_menu,
    !.

/*****/

/***** FBS SECOND RESULT *****/
check( _, _, [111|_] ):-
    assert(yes(111)),
    clearwindow,
    write("\n\tPerform another FBS and enter result"),
    fbs_menu_2,
    !.

/*****/

/***** TRACE ASSESS 2 *****/
/* when > 0.04 u/hr is true */

check( _, _, [133|_] ):-
    assert(yes(133)),
    clearwindow,
    write("\nObserve trace for 30 minutes:"),
    trace_condition_menu_2,
    !.

/*****/

/***** TRACE ASSESS 2 *****/
/* when < 0.04 u/hr is true */

check( _, _, [134|_] ):-
    assert(yes(134)),
    clearwindow,
    write("\nObserve trace for 30 minutes:"),
    trace_condition_menu_2,
    !.

/*****/

/***** TRACE ASSESS 3 *****/
/* when first FBS is 7.20 - 7.24 */

check( _, _, [132|_] ):-
    assert(yes(132)),
    clearwindow,

```

```

        write("\nObserve trace until one of the following conditions
becomes true:"),
        trace_condition_menu_3,
        !.
/*****
/***** FBS SECOND RESULT *****/
/* when first result was 7.20 - 7.24 */
check( _, _, [129|_] ):-
    assert(yes(129)),
    clearwindow,
    write("\n\tPerform another FBS and enter result"),
    fbs_menu_2,
    !.
/*****
/***** TRACE ASSESS 4 *****/
/* when first FBS 7.20-7.24 & second FBS was 7.20-7.24 */
check( _, _, [130|_] ):-
    assert(yes(130)),
    clearwindow,
    write("\nObserve trace for 30 minutes:"),
    trace_condition_menu_2,
    !.
/*****

check( RNO, HISTORY, [BNO|REST] ):- yes(BNO), !,
    check(RNO, HISTORY, REST).

check( _, _, [BNO|_] ):- no(BNO), !, fail.

check( RNO, HISTORY, [BNO|REST] ):- cond(BNO,NCOND),
    fronttoken(NCOND,"not",_COND),
    frontchar(_COND,_COND),
    cond(BNO1,COND),
    notest(BNO1), !,
    check(RNO, HISTORY, REST),
    write_data(BNO).

check( _, _, [BNO|_] ):- cond(BNO,NCOND),
    fronttoken(NCOND,"not",_COND),
    frontchar(_COND,_COND),
    cond(BNO1,COND),
    yes(BNO1),
    !, fail.

check( RNO, HISTORY, [BNO|REST] ):-
    cond(BNO,TEXT),
    inpq(HISTORY,RNO,BNO,TEXT),
    check(RNO, HISTORY, REST).

    check( _, _, [] ).

notest(BNO):-no(BNO),!.
notest(BNO):-not(yes(BNO)),!.
inpq(HISTORY,RNO,BNO,TEXT):-
    clearwindow,
    write("\n\n"),
    write(TEXT,": "),

```

```

    ROW = 6,
    COL = 70,
    menu(ROW, COL, 15, 9, [yes, no], "", 1, CHOICE),
    do_answer(HISTORY, RNO, TEXT, BNO, CHOICE).

do_answer(, , , BNO, 1) :-assert(yes(BNO)),
write_data(BNO).

do_answer(, , , BNO, 2) :-assert(no(BNO)),
    shiftwindow(1), write(no), nl, fail.

/*Norm CTG*/
norm_ctg:-clearwindow,
write("\n\n\tNormal CTG"),
writedevicemyfile),
write("0\n"),
write("1\n"),
writedevicem(screen).

/*Input info*/

/* Enter antenatal risk factors */
antenatal_risk_info:-
    assert(risk_factor("present")),
    risk_factor(X),
    assert_risk_cond(X),
    write_data(0),
    write_data(7),
    repeat,
        shiftwindow(1),
        menu(6, 20, 71, 12,
            ["confirmed IUGR",
            "abnormal antenatal trace",
            "abruption antenatally",
            "none",
            "Exit"],
            "ENTER ANTENATAL RISK FACTORS",
            2,
            OPTION),
        choice(OPTION),
        stop(OPTION),
        !.

choice(0):-write_data(0).
choice(1):-write_data(1).
choice(2):-write_data(2).
choice(3):-write_data(3).
choice(4):-write_data(4).
choice(5):-write_data(0).

stop(0):-clearwindow.
stop(4):-clearwindow,
    retract(yes(26)),
    retractall(risk_factor(_)),
    assert(risk_factor("none")),
    risk_factor(X),
    assert_risk_cond(X).

stop(5):-clearwindow.
/***** LABOUR RISK FACTORS *****/

```

```

labour_risk_info:-
  clearwindow,
  write_data(0),
  write_data(3),
  repeat,
    shiftwindow(1),
    menu(6,20,71,12,
      ["Thick meconium liquor",
      "Prolonged deceleration > 8 mins",
      "Exit"],
      "UPDATE LABOUR RISK FACTORS",
      2,
      OPTION),
    choi(OPTION),
    sto(OPTION),
    !.

choi(0):-write_data(0).
choi(1):-write_data(1),
  add_risk.

choi(2):-write_data(2),
  add_risk.

choi(3):-write_data(0).
sto(0):-clearwindow.
sto(3):-clearwindow.

add_risk:-
  retractall(risk_factor(_)),
  retractall(no(_)),
  retractall(yes(_)),
  assert(risk_factor("present")),
  risk_factor(X),
  assert_risk_cond(X).

/***** MENUS FOR TRACE INTERPRETATION *****/

baseline_assess:-
  repeat,
    shiftwindow(1),
    menu(8,25,14,15,
      ["Less than 90 for > 3 minutes",
      "Between 90 - 110",
      "Between 110 - 160",
      "Between 160 - 180",
      "Greater than 180"],

      "BASELINE",2,CHOICE),
    proces1(CHOICE),
    stop1(CHOICE),!.

  proces1(1):-assert(yes(1)), assert(no(13)), assert(no(28)),
assert(no(29)), assert(no(30)).
  proces1(2):-assert(no(1)), assert(yes(13)), assert(no(28)),
assert(no(29)), assert(no(30)).
  proces1(3):-assert(no(1)), assert(no(13)), assert(no(28)),
assert(no(29)), assert(yes(30)).
  proces1(4):-assert(no(1)), assert(no(13)), assert(yes(28)),
assert(no(29)), assert(no(30)).

```

```
proceso1(5):-assert(no(1)), assert(no(13)), assert(no(28)),
assert(yes(29)), assert(no(30)).
```

```
stop1(_):-clearwindow.
```

```
variability_assess:-
```

```
repeat,
    shiftwindow(1),
    menu(8,30,14,15,
        ["Absent",
         "Reduced (< 5 bpm)",
         "Normal",
         "Increased (> 25 bpm)"],
        "VARIABILITY",2,CHOICE),
    proceso2(CHOICE),
stop2(CHOICE),!.
```

```
proceso2(1):-assert(no(14)), assert(no(15)), assert(no(16)),
assert(yes(78)).
```

```
proceso2(2):-assert(yes(14)), assert(no(15)), assert(no(16)),
assert(no(78)).
```

```
proceso2(3):-assert(no(14)), assert(yes(15)), assert(no(16)),
assert(no(78)).
```

```
proceso2(4):-assert(no(14)), assert(no(15)), assert(yes(16)),
assert(no(78)).
```

```
stop2(_):-clearwindow.
```

```
accelerations_assess:-
```

```
repeat,
    shiftwindow(1),
    menu(8,30,14,15,
        ["Absent",
         "Present"],
        "ACCELERATIONS",2,CHOICE),
    proceso3(CHOICE),
stop3(CHOICE),!.
```

```
proceso3(1):-assert(no(20)).
```

```
proceso3(2):-assert(yes(20)).
```

```
stop3(_):-clearwindow.
```

```
decelerations_assess:-
```

```
repeat,
    shiftwindow(1),
    menu(8,30,14,15,
        ["None",
         "Early or Mild Variable",
         "Late or Severe Variable"],
        "DECELERATIONS",2,CHOICE),
    proceso4(CHOICE),
stop4(CHOICE),!.
```

```
proceso4(1):-assert(yes(17)), assert(no(18)), assert(no(19)).
```

```
proceso4(2):-assert(no(17)), assert(yes(18)), assert(no(19)).
```

```
proceso4(3):-assert(no(17)), assert(no(18)), assert(yes(19)).
```

```
stop4(_):-clearwindow.
```

```
fbs_menu:-
```

```
repeat,
    shiftwindow(1),
```

```

    menu(8,30,14,15,
        ["Greater or equal to 7.25",
        "Between 7.20 and 7.24",
        "Less than 7.20"],
        "FBS pH",2,CHOICE),
        proces5(CHOICE),
stop5(CHOICE),!.

proces5(1):-assert(yes(101)), assert(no(102)), assert(no(103)).
proces5(2):-assert(no(101)), assert(yes(102)), assert(no(103)).
proces5(3):-assert(no(101)), assert(no(102)), assert(yes(103)).
stop5(_):-clearwindow.

fbs_menu_2:-
repeat,
    shiftwindow(1),
    menu(8,30,14,15,
        ["Greater or equal to 7.25",
        "Between 7.20 and 7.24",
        "Less than 7.20"],
        "FBS pH",2,CHOICE),
        proces6(CHOICE),
stop6(CHOICE),!.

proces6(1):-assert(yes(112)), assert(no(113)), assert(no(114)).
proces6(2):-assert(no(112)), assert(yes(113)), assert(no(114)).
proces6(3):-assert(no(112)), assert(no(113)), assert(yes(114)).
stop6(_):-clearwindow.

trace_condition_menu:-
repeat,
    shiftwindow(1),
    menu(8,30,14,15,
        ["Returns to Acceptible Trace",
        "Trace Deteriorates",
        "60 minutes elapse since last FBS"],
        "TRACE ASSESSMENT",2,CHOICE),
        proces7(CHOICE),
stop7(CHOICE),!.

proces7(1):-assert(yes(105)), assert(no(106)), assert(no(107)).
proces7(2):-assert(no(105)), assert(yes(106)), assert(no(107)).
proces7(3):-assert(no(105)), assert(no(106)), assert(yes(107)).
stop7(_):-clearwindow.

trace_condition_menu_2:-
repeat,
    shiftwindow(1),
    menu(8,30,14,15,
        ["Returns to an Acceptible Trace",
        "Trace remains the same",
        "Trace Deteriorates"],
        "TRACE ASSESSMENT",2,CHOICE),
        proces8(CHOICE),
stop8(CHOICE),!.

proces8(1):-assert(yes(117)), assert(no(118)), assert(no(119)).
proces8(2):-assert(no(117)), assert(yes(118)), assert(no(119)).
proces8(3):-assert(no(117)), assert(no(118)), assert(yes(119)).
stop8(_):-clearwindow.

```



```

trace_condition_menu_3:-
  repeat,
    shiftwindow(1),
    menu(10,6,14,15,
      ["Returns to an Acceptible Trace and is maintained for 60
minutes",
      "Trace Persisting for a total of 30 minutes",
      "Trace Deteriorates",
      "Trace previously severely abnormal and this re-occurs"],
      "TRACE ASSESSMENT",2,CHOICE),
    proces9(CHOICE),
    stop9(CHOICE),!.

  proces9(1):-assert(yes(126)), assert(no(127)), assert(no(128)),
assert(no(131)).
  proces9(2):-assert(no(126)), assert(yes(127)), assert(no(128)),
assert(no(131)).
  proces9(3):-assert(no(126)), assert(no(127)), assert(yes(128)),
assert(no(131)).
  proces9(4):-assert(no(126)), assert(no(127)), assert(yes(128)),
assert(yes(131)).
  stop9(_):-clearwindow.

v_absent_d_nolates_nopeth_menu:-
  repeat,
    shiftwindow(1),
    menu(10,25,14,15,
      ["Returns to an Acceptible Trace",
      "Trace Persisting for 30 minutes",
      "Trace Deteriorates"],
      "TRACE ASSESSMENT",2,CHOICE),
    proces10(CHOICE),
    stop10(CHOICE),!.

  proces10(1):-assert(yes(75)), assert(no(76)), assert(no(77)).
  proces10(2):-assert(no(75)), assert(yes(76)), assert(no(77)).
  proces10(3):-assert(no(75)), assert(no(76)), assert(yes(77)).
  stop10(_):-clearwindow.

v_absent_d_nolates_peth_menu:-
  repeat,
    shiftwindow(1),
    menu(10,25,14,15,
      ["Returns to an Acceptible Trace",
      "Trace Persisting for 60 minutes",
      "Trace Deteriorates"],
      "TRACE ASSESSMENT",2,CHOICE),
    proces11(CHOICE),
    stop11(CHOICE),!.

  proces11(1):-assert(yes(75)), assert(no(76)), assert(no(77)).
  proces11(2):-assert(no(75)), assert(yes(76)), assert(no(77)).
  proces11(3):-assert(no(75)), assert(no(76)), assert(yes(77)).
  stop11(_):-clearwindow.

v_absent_d_lates_menu:-
  repeat,
    shiftwindow(1),

```

```

    menu(10,25,14,15,
        ["Returns to an Acceptible Trace",
        "Trace Persisting for 15 minutes",
        "Trace Deteriorates"],
        "TRACE ASSESSMENT",2,CHOICE),
        proces12(CHOICE),
stop12(CHOICE),!.

proces12(1):-assert(yes(75)), assert(no(76)), assert(no(77)).
proces12(2):-assert(no(75)), assert(yes(76)), assert(no(77)).
proces12(3):-assert(no(75)), assert(no(76)), assert(yes(77)).
stop12(_):-clearwindow.

cervical_dilatation_menu:-
repeat,
    shiftwindow(1),
    menu(10,25,14,15,
        ["Less than 8 cm",
        "Between 8 and 10 cm",
        "In ACTIVE second stage of labour"],
        "CERVICAL DILATATION",2,CHOICE),
        proces13(CHOICE),
stop13(CHOICE),!.

proces13(1):-assert(yes(43)), assert(no(44)), assert(no(45)).
proces13(2):-assert(no(43)), assert(yes(44)), assert(no(45)).
proces13(3):-assert(no(43)), assert(no(44)), assert(yes(45)).
stop13(_):-clearwindow.

/*****
*/

/*User commands*/

write_data(X):-writedevicemyfile(myfile),
write(X, "\n"),
writedevicemyfile(screen).

load_know:-consult("infant.gni").

erase:-retract(_),fail.
erase.

clear:-retractall(yes(_)),retractall(no(_)).

/*system commands*/

reverse(X,Y):-
    reversel([],X,Y).
reversel(Y,[],Y).
reversel(X1,[U|X2],Y):-reversel([U|X1],X2,Y).

```

Appendix G

An inference engine written in 'C' software.

```
/* An expert system written in C for the */
/* management of labour. */
```

```
/* Robert D. F. Keith - 1992 */
```

```
#include <stdio.h>
#include <stdlib.h>
#include <graphics.h>
#include <math.h>
#include <string.h>
#include <alloc.h>
#include <conio.h>

/*** libraries to include ****/

#define CLS printf("%c[2J",27)
#define TRUE 1
#define FALSE 0

typedef struct
{
    int bl;
    char hrvar;
    int noaccs;
    int dipear;
    int diplat;
    int dipearsev;
    int diplatsev;
    int nocons;
    int tnoise;
    int snoise;
    char clas;
} feats;

typedef struct {
    char rule[55];
} kb_rule_struct;

typedef struct {
    char condition[55];
} kb_cond_struct;

typedef struct
{
    int fbs_time;
    float fbs_ph;
} fbs_res;

int numb_of_fbs;
int get_rule();
extern class_ctg();
extern display_conditions();
extern special_condition1();
extern special_condition2();

char current_head[100];
char statement_type[20], rule_number[10], head[30], tail[300];

int cond[20], number_conds, cond_ptr, status;
int yes[100], no[100];
int number_in_no, number_in_yes;
int yes_found_flag = FALSE, no_found_flag = FALSE;
```

```

int yes_found_flag = FALSE, no_found_flag = FALSE;
kb_rule_struct kb_r[300];
kb_cond_struct kb_c[150];
fbs_res fbs_res_array[10];

int rule_position;
int max_rule;
int inhibit_flag;

int validation_score; // used for validation to measure concern

call_expert(feats segt[], int time)
{
    if(time == 15)          /* time == 16 */
    {
        load_kb();        /* only done once - stays in memory */
        set_yes_no();     /* reset when a goal is reached */
        setfillstyle(1,9);
        bar(0,310,640,480);
    }

    display_extracted_features(segt, time);
    class_ctg(segt, time, yes, no, inhibit_flag);
    if(inhibit_flag != TRUE)
    {
        number_in_no = 10;
        number_in_yes = 4;
    }

    rule_position = 0;
    validation_score = 0;
    inference_engine(segt, time);

    prn_validation_score();
    //getch();
    if(validation_score == 5)
        exit(0);
}

/*****
/***** INFERENCE ENGINE *****/
inference_engine(feats segt[], int time)
{
    int flag, goal_flag = TRUE;
    int j, k;

    strcpy(current_head, "quality");

    while(TRUE)
    {
        no_found_flag = FALSE;

        /* DOES CURRENT HEAD EXIST IN KB. IF IT DOES BREAK - *****/
        /* IF IT DOESN'T, IT MUST BE THE GOAL SOLUTION ELSE NO SOLUTION EXISTS */

        do
        {
            flag = get_rule(); /* GET_RULE() WILL ALSO SPLIT THE RULE */
            if(flag != 0)
                {

```

```

        if(goal_flag == TRUE)
            {
                goal_message(current_head);
                set_yes_no();
                return;
            }
        if(goal_flag == FALSE)
            {
                printf("\n\nI could not find a solution\n");
                getch();
                set_yes_no();
                return;
            }
    }
    rule_position++;
}
while(strcmp(current_head, head) != 0);
    /***/

goal_flag = FALSE; /* a rule with the current head has been found */
    /* therefore if a complete solution does not exist */
    /* it must be an error. Goal_flag is only set TRUE */
    /* when a rule is satisfied */

/***/ CHECK WHETHER ANY CONDITIONS ASSOCIATED WITH THE CURRENT RULE ***/
/***/ EXIST IN THE NO[] ARRAY - IF THEY DO THEN THE CURRENT RULE ***/
/***/ CANNOT BE TRUE ***/

for(j = 0; j < number_conds; j++)
    {
        for(k = 0; k < number_in_no; k++)
            {
                if(cond[j] == no[k])
                    no_found_flag = TRUE;
            }
    }
    /***/

cond_ptr = 0; /* begin with first condition */

while(no_found_flag == FALSE)
    {
        yes_found_flag = FALSE;

        /*if all the conditions associated with the current rule are true then */
        /* the rule must be true in which case the tail belonging to the rule */
        /* becomes the current head are possibly a goal state */

        if(cond_ptr >= number_conds)
            {
                look_up(); // look up table for validation concern
                display_conditions(cond, number_conds, rule_number);
                strcpy(current_head, tail);
                goal_flag = TRUE;
                rule_position = 0;
                break;
            }
            /***/

        /* test to see whether the current condition is already known to be true */
        for(k = 0; k < number_in_yes; k++)

```

```

        {
            if(cond[cond_ptr] == yes[k])
            {
                yes_found_flag = TRUE;
                cond_ptr++;
            }
        }

/* if it is then continue with the next condition */

if(yes_found_flag == TRUE)
    continue;

/****** CHECK COND *****/
/* if the answer to the current condition is unknown then we must try and
   prove the condition */
    rule_position = 0;

/* check for "not" & special condition */

        if(check_condition(cond[cond_ptr], segt, time) == TRUE)
            return;
    cond_ptr++; /* next condition */
        /****** */
    }
}

/******
***** LOAD KB *****/
load_kb()
{
FILE *know_file;
char entry[50], entry_test[50];
int rule_array_pos, cond_array_pos, j;
char *ptr;

if((know_file = fopen("inf.kb", "r")) == NULL)
    {
        printf("cannot open file");
        getch();
        exit(0);
    }

rule_array_pos = 0;
cond_array_pos = 0;

while(feof(know_file) == 0)
    {
        fgets(entry, 50, know_file);
        strcpy(entry_test, entry);
        ptr = strtok(entry_test, "(");
        if(strcmp(ptr, "rule") == 0)
            {
                strcpy(kb_r[rule_array_pos].rule, entry);
                rule_array_pos++;
            }
        if(strcmp(ptr, "cond") == 0)
            {
                strcpy(kb_c[cond_array_pos].condition, entry);

```

```

        cond_array_pos++;
    }
}
max_rule = rule_array_pos;
}
/*****
/***** GET RULE *****/
int get_rule()
{
int test, dummy, j;
char *ptr, current_rule[100];

number_conds = 0;
for(j = 0; j < 20; j++)
    cond[j] = 0;
if(rule_position > max_rule)
    return(1);
strcpy(current_rule, kb_r[rule_position].rule);

ptr = strtok (current_rule, "(),");

ptr = strtok (NULL, "(),");
strcpy(rule_number, ptr);      /* get rule number */

ptr = strtok (NULL, "(),");   /* get rule head */
strcpy(head, ptr);

                                /* get rule tail */
ptr = strtok (NULL, "(),");
strcpy(tail, ptr);

while(TRUE)
    {
        ptr = strtok (NULL, "(),");   /* get the conditions associated
with the rule */
        if(ptr == NULL)
            break;
        dummy = atoi(ptr);
        if(dummy > 0)                /* valid conditions > 0 */
            {
                cond[number_conds] = dummy;
                number_conds++;
            }
    }

return(0);
}
/*****
/***** SET YES NO ARRAYS TO 0 *****/
set_yes_no()
{
int j;

for(j = 0; j < 100; j++)
    {
        yes[j] = 0;
        no[j] = 0;
    }
number_in_yes = 0;
number_in_no = 0;

```



```

number_conds = 0;
inhibit_flag = FALSE;
}

/*****

/***** CHECK CONDITION *****/
check_condition(int condition_number, feats segt[], int time)
{
char cond_str[10], cond_search[50];
char statement[500], question[500];
char *ptr;
int j, c_ptr;

    /* find the condition given by the token cond(x, */
    /* where x is the condition number */

c_ptr=0;
do
    {
    itoa(condition_number, cond_str, 10);
    strcpy(cond_search, "cond(");
    strcat(cond_search, cond_str);
    strcpy(statement, kb_c[c_ptr].condition);
    c_ptr++;
    }
while(strstr(statement, cond_search) == NULL);

ptr = strtok (statement, ",");

ptr = strtok (NULL, ")");
strcpy(question, ptr);

        /* NOT CONDITIONS */
/*a "not" condition must always follow numerically after the condition */
/* for which the "not" condition is the opposite response */
/* therefore if the condition number is 22 then the "not" condition */
/* if it exists, must be condition number 23 */

if(strncmp(question, "not", 3) == 0)
        /*checks first three chars of question */
    {
    /* check if the condition has been asked before and the response */
    /* was no, in which case the "not" condition will be yes. This */
    /* will virtually always be the case as long as the KB has been */
    /* constructed with care */

        for(j = 0; j < number_in_no; j++)
            if(no[j] == condition_number - 1)
                {
                yes[number_in_yes] = condition_number;
                number_in_yes++;
                return(FALSE);
                }

    /* check if the condition has been asked before and the response */
    /* was yes, in which case the "not" condition will no. This will */
    /* seldom be the case */

        for(j = 0; j < number_in_yes; j++)
            if(yes[j] == condition_number - 1)

```

```

        {
        no[number_in_no] = condition_number;
        number_in_no++;
        no_found_flag = TRUE;
        /* fsetpos(fp, &filepos); */
        return(FALSE);
        }
    }

    /*****/

if(condition_number < 500)
    status = special_condition1(condition_number, segt, time,
&number_in_yes, &number_in_no, yes, no, fbs_res_array);
if(condition_number >= 500)
    status = special_condition2(condition_number, segt, time,
&number_in_yes, &number_in_no, yes, no, fbs_res_array);

if(status != 0)
    {
    if(status == 1)
        {
        no[number_in_no] = condition_number;
        number_in_no++;
        no_found_flag = TRUE;
        }
    if(status == 2)
        {
        yes[number_in_yes] = condition_number;
        number_in_yes++;
        yes_found_flag = TRUE;
        }
    if(status == 3)
        {
        yes[number_in_yes] = condition_number;
        number_in_yes++;
        yes_found_flag = TRUE;
        no[number_in_no] = condition_number+1;
        number_in_no++;
        }
    if(status == 4)
        {
        no[number_in_no] = condition_number;
        number_in_no++;
        no_found_flag = TRUE;
        yes[number_in_yes] = condition_number+1;
        number_in_yes++;
        }
    if(status == 5)
        {
        inhibit_flag = TRUE;
        ask(condition_number, 0);
/* display possible message, file must exist in "question" directory.
response_flag = 0 (FALSE) */

        return(TRUE); /* keep positon in knowledge tree */
        }
        ask(condition_number, 0);
/* display possible message, file must exist in "question" directory.
response_flag = 0 (FALSE) */

        return(FALSE);
    }

```

```

    }
if(status == 0)
    ask(condition_number, 1); /* response_flag = 1 TRUE. A question is
to be asked, a yes/no
                                response is required */
    }
/*****
/***** ASK *****/
ask(int condition_number, int response_flag)
{
/*    response_flag = TRUE :- a question is being asked requiring a yes/no
response response_flag = FALSE :- a message is displayed requiring only
an acknowledgment */

#define          TLX    40
#define          TLY    199
#define          BLX    600
#define          BLY    290

FILE *gm;
int dot, j, k, loop;
char *ptr;
void far *buf;
unsigned size;
char out_msg[200];
char ans, cond_num_str[10], quest_file[30];

itoa(condition_number, cond_num_str, 10);
strcpy(quest_file, "question\\");
strcat(quest_file, cond_num_str);

if((gm = fopen(quest_file, "rt")) == NULL)
    {
        if(response_flag == TRUE)
            {
                printf("I could not think of the correct question to ask\n");
                printf("The condition I was looking for was %d",
condition_number);
                getch();
                exit(0);
            }
        else
            return;
    }
fgets(out_msg, 500, gm);
fclose(gm);

size = imagesize(TLX, TLY, BLX, BLY); /* max 108k */
if((buf = farmalloc(size)) == NULL)
    {
        printf("not enough memory");
        getch();
        exit(0);
    }
getimage(TLX, TLY, BLX, BLY, buf);

if(response_flag == TRUE)
    {
        set_up_screen(TLX, TLY, BLX, BLY, 8);
        outtextxy(((BLX-TLX)/2)-32, TLY+10, "QUESTION");
    }

```

```

    }
if(response_flag == FALSE)
    {
    set_up_screen(TLX,TLY,BLX,BLY,4);
    outtextxy(((BLX-TLX)/2)-32, TLY+10,"MESSAGE");
    }

ptr = strtok(out_msg, "|!");
loop = 0;
while(TRUE)
    {
    outtextxy(TLX+15, (TLY+30)+15*loop,ptr);
    if((strrchr(ptr, '.')) != NULL)
        break;

    ptr = strtok(NULL, "|!");
    loop++;
    }

while(response_flag == TRUE)
    {

    ans = getch();          /** A CHANGE **/
    if(ans == 'y')
        {
        outtextxy(TLX+15, TLY+65, "YES");
        yes[number_in_yes] = condition_number;
        number_in_yes++;
        break;
        }
    if(ans == 'n')
        {
        no[number_in_no] = condition_number;
        number_in_no++;
        no_found_flag = TRUE;
        break;
        }
    }

if(response_flag == FALSE)
    // sleep(3);
    getch();

putimage(TLX,TLY, buf, COPY_PUT);
farfree(buf);

}
/*****

/***** DISPLAY EXTRACTED FEATURES *****/
display_extracted_features(feats segt[], int time)
{
char bline_str[10], hrvar_str[20], noaccs_str[10];
char ear_str[10], lat_str[10], earsev_str[10], latsev_str[10];
char cons_str[10], time_str[25], hours_str[10], minutes_str[10];
int minutes, hours;

itoa(segt[time/5 - 3].bl, bline_str, 10);

if(segt[time/5 - 3].hrvar == 'a')
    strcpy(hrvar_str, "Absent");
if(segt[time/5 - 3].hrvar == 'r')

```

```

        strcpy(hrvar_str, "Reduced");
if(segtime[time/5 - 3].hrvar == 'n')
        strcpy(hrvar_str, "Normal");
if(segtime[time/5 - 3].hrvar == 'i')
        strcpy(hrvar_str, "Increased");

itoa(segtime[time/5 - 3].noaccs, noaccs_str, 10);
itoa(segtime[time/5 - 3].dipear, ear_str, 10);
itoa(segtime[time/5 - 3].diplat, lat_str, 10);
itoa(segtime[time/5 - 3].dipearsev, earsev_str, 10);
itoa(segtime[time/5 - 3].diplatsev, latsev_str, 10);
itoa(segtime[time/5 - 3].nocons, cons_str, 10);

minutes = 0;
hours = 0;

while(TRUE)
    {
        if(time - 60 >=0)
            {
                hours++;
                time -= 60;
            }

        if(time - 60 < 0)
            {
                minutes = time;
                break;
            }
    }
itoa(hours, hours_str, 10);
itoa(minutes, minutes_str, 10);

strcpy(time_str, hours_str);
strcat(time_str, " h:");
strcat(time_str, minutes_str);
strcat(time_str, " m");

setfillstyle(1,9);
bar(0,200, 210, 300);
setcolor(WHITE);
rectangle(1,201,209,299);
rectangle(3,203,207,297);
setcolor(WHITE);
outtextxy(7, 207, "Baseline:");
outtextxy(132, 207, bline_str);
outtextxy(7, 217, "Variability:");
outtextxy(132, 217, hrvar_str);
outtextxy(7, 227, "No. Accelns:");
outtextxy(132, 227, noaccs_str);
outtextxy(7,237, "Earlies:");
outtextxy(132, 237, ear_str);
outtextxy(7,247, "Lates:");
outtextxy(132, 247, lat_str);
outtextxy(7, 257, "Severe Earlies:");
outtextxy(132, 257, earsev_str);
outtextxy(7, 267, "Severe Lates:");
outtextxy(132, 267, latsev_str);
outtextxy(7, 277, "Contractions:");
outtextxy(132, 277, cons_str);
outtextxy(7, 287, "Time elapsed:");

```

```

outtextxy(132, 287, time_str);
}
/*****
/***** GOAL MESSAGE *****/
goal_message(char goal_file[])
{

#define          TLX    40
#define          TLY    199
#define          BLX    600
#define          BLY    290

FILE *gm;
int dot, j, k, loop;
char *ptr;
void far *buf;
unsigned size;
char out_msg[500];

gm = fopen(goal_file, "rt");
fgets(out_msg, 500, gm);
fclose(gm);

size = imagesize(TLX,TLY,BLX,BLY);    /* max 108k */
if((buf = farmalloc(size)) == NULL)
    {
        printf("not enough memory");
        getch();
        exit(0);
    }
getimage(TLX,TLY,BLX,BLY, buf);
set_up_screen(TLX,TLY,BLX,BLY,4);

ptr = strtok(out_msg, "|!");
loop = 0;
while(TRUE)
    {
        outtextxy(TLX+10, (TLY+10)+15*loop,ptr);
        if((strchr(ptr, '.')) != NULL)
            break;

        ptr = strtok(NULL, "|!");
        loop++;
    }

//sleep(3);
getch();
putimage(TLX,TLY, buf, COPY_PUT);
free(buf);

}
/*****
/***** SET UP SCREEN *****/
set_up_screen(int x1, int y1, int x2, int y2, int bkg_col)
{
setfillstyle(1,bkg_col);
bar(x1,y1,x2,y2);
setcolor(WHITE);
rectangle(x1+2,y1+2,x2-2,y2-2);

```

```
rectangle(x1+4,y1+4,x2-4,y2-4);
}
/*****/

/***** PRINT VALIDATION SCORE *****/

prn_validation_score()
{
char validation_score_str[5];

setfillstyle(1,15);
bar(100,70, 200, 115);
itoa(validation_score, validation_score_str, 10);
setcolor(0);
outtextxy(109, 75, "Validation");
outtextxy(130, 85, "Score");
outtextxy(145, 100,validation_score_str);
}
/*****/
```

Appendix H

Case information and results obtained in the validation of the system.

For each of the 50 cases considered in the validation study (chapter 5), the following are given;

1. Previous obstetric history.
2. Relevant labour events.
3. Perinatal outcome.
4. Graph of cervical dilatation and estimated fetal scalp blood pH.
5. The recorded scores from the two reviews (1 and 2) recorded by the experts (A - Q) and system (S).
6. The calculated agreement between each pair of review sequences.

Case 1 1289

Mrs M.C. is a primigravida. She is a fit lady and does not smoke. Her pregnancy progressed normally and she laboured spontaneously at 39 weeks gestation, being admitted on 5.4.91 at 01.00hrs. At this time the fetal head was 3/5 palpable and the cervix part effaced but closed. Pethidine 100mg was given at 05.00hrs.

Labour events

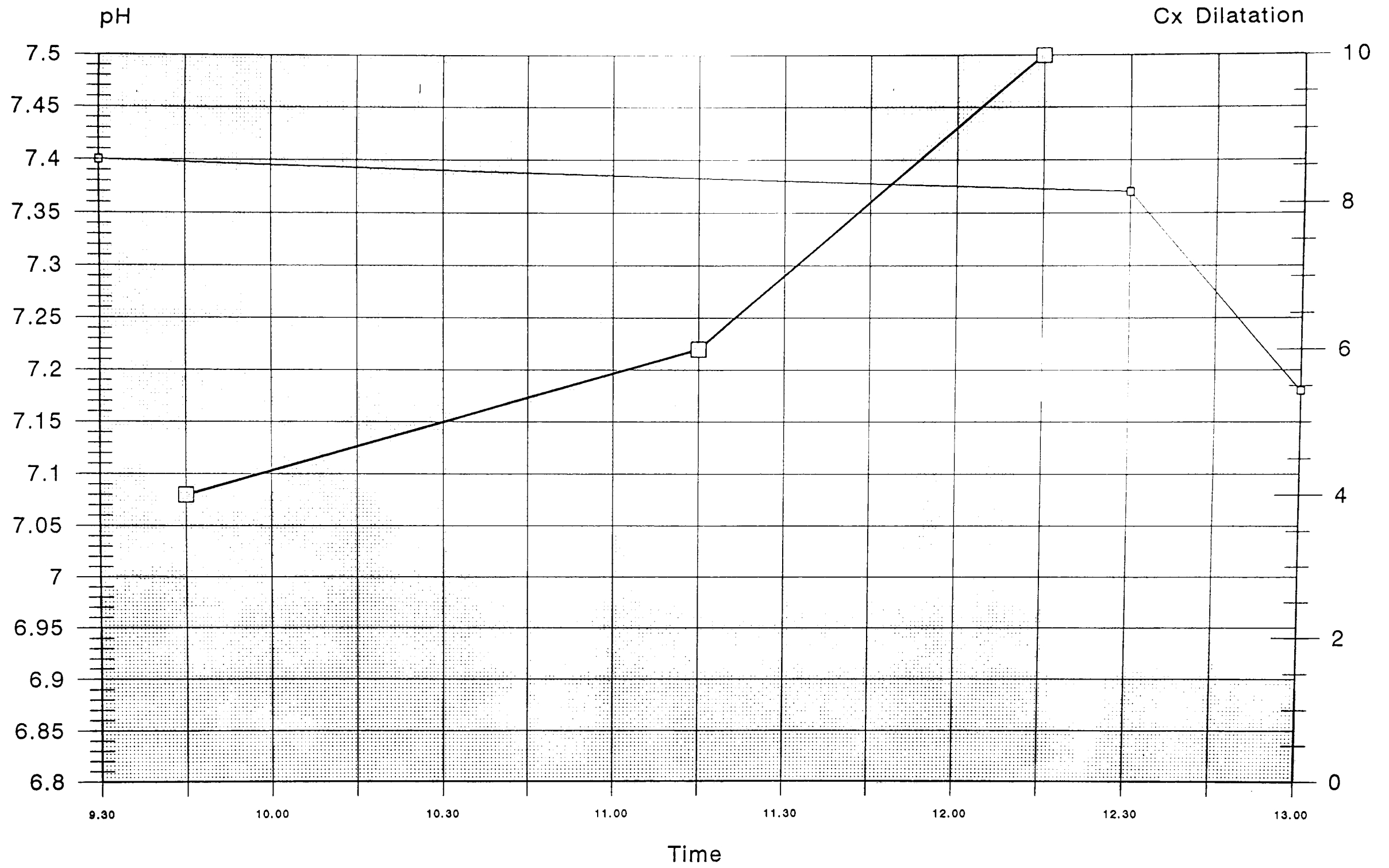
- 09.50 VE Cx 4cm dilated, effaced. Station -1. Direct OP position. ARM; clear liquor. FSE applied.
- 10.15 Epidural begun.
- 10.30 Epidural working.
- 11.10 VE Cx 6cm dilated, thin. Station 0.
- 11.50 Top up.
- 12.15 VE Cx fully dilated. Station +1. LOT position. Caput + moulding +.
- 12.25 FBS; pH 7.37 BE -3
- 12.59 Female infant delivered by Kjellands forceps. Indication; prolonged FH decelerations.

Outcome

Birth weight 3.38kg

Apgar 9 & 9

Cord gases pH 7.18 / 7.29 BD(ecf) 6 / 4



□ pH □ Cx Dilatation (cm)

Case 2 1343

Mrs J.R. is a 31 year old lady expecting her first baby. She is fit and well, and does not smoke. Her pregnancy was straightforward until 38 weeks, when she developed mild hypertension (BP 130/90) without proteinuria. BP remained stable and fetal size was judged to be appropriate for dates. Labour was induced at 40 weeks because of persistent mild hypertension and maternal insistence. On admission at 11.20hrs on 11.4.91, presentation was cephalic 3/5, and cervix closed, uneffaced and firm. Prostin gel 2mg was given PV, and a further 2mg dose at 19.15hrs (Bishop's score had not improved).

Labour events

Note; maternal BP very stable at 130/70 to 140/80 throughout labour.

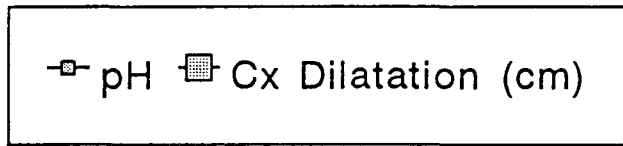
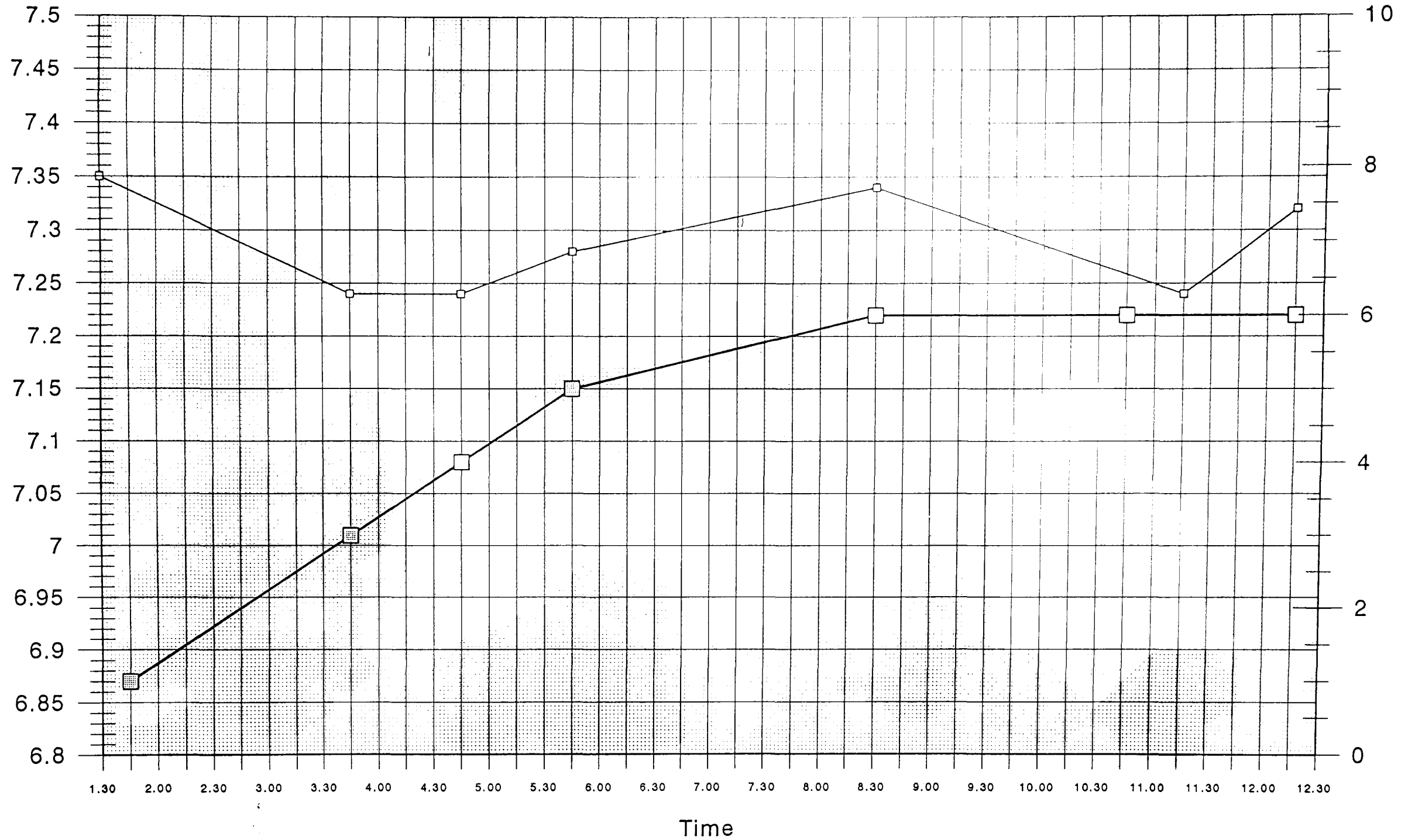
- 23.00 Contractions becoming more regular.
- 00.50 Pethidine 100mg, Phenergan 25mg im.
- 01.50 Short but very frequent contractions. VE Cx 1cm dilated, effaced and thin. ARM; clear liquor. FSE applied.
- 02.00 Ranitidine 50mg, metoclopramide 10mg iv.
- 02.30 Epidural inserted.
- 03.50 VE Cx 3cm dilated, thin. Station -2.
FBS; pH 7.24 BE -8
- 04.40 VE Cx 4cm dilated. Station -2.
FBS; pH 7.24 BE -9 Maternal sample; pH 7.44
- 05.12 Top up.
- 05.45 VE Cx 5cm dilated. Station -1.
FBS; pH 7.28 BE -5
- 07.40 Top up.
- 08.30 VE Cx 6cm dilated. Station -1. Caput + No moulding.
FBS; pH 7.34 BE -4
- 08.40 Syntocinon started.
- 09.15 Top up.
- 10.45 VE Cx 6cm dilated, now thick & oedematous. Station -1. Moulding +.
- 11.00 Top up.
- 11.20 FBS; pH 7.24 BE -6 Decision for emergency Caesarean section.
- 12.20 Male infant delivered by C/S under epidural. Indication; falling fetal scalp pH and poor progress in labour.

Outcome

Birth weight 3.96kg
Apgar 9 & 9
Cord gases pH 7.32 / 7.36 BD(ecf) 1 / 2

pH

Cx Dilatation



Case 3 29

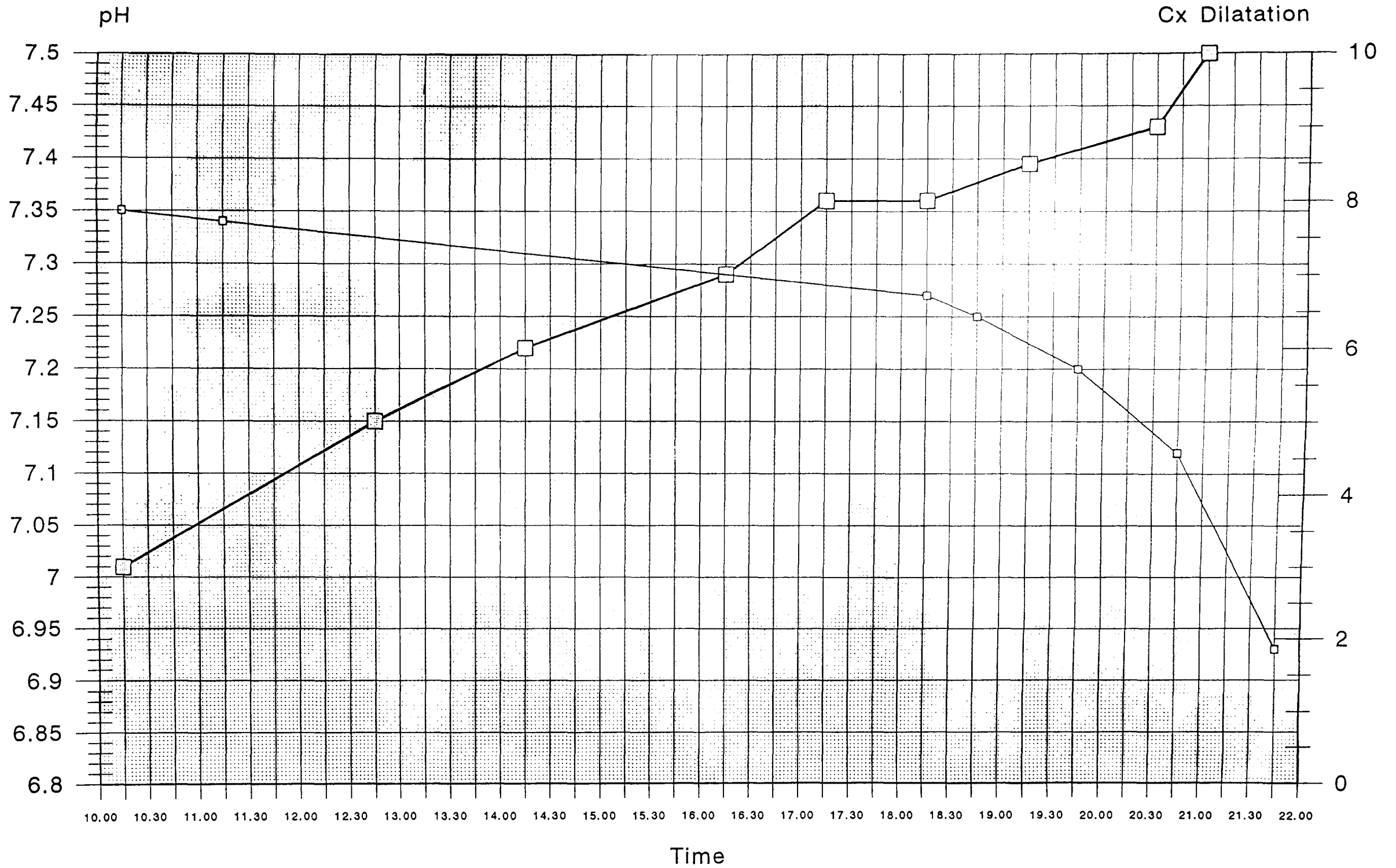
Mrs G.G. is a 29 year old lady expecting her second baby. She has a five year old boy who was delivered normally at term, weighing 9lb 2oz. The current pregnancy was straightforward, although labour was induced at 42 weeks on the grounds of postmaturity. It was suspected at this stage that the baby was fairly large for dates. Mrs G.G. has no significant medical history and she does not smoke. On admission on 11.2.92 the presentation was cephalic; cervix soft but closed and uneffaced. 3 doses of Prostin (3mg tablets) were given over the next 48 hours. At 09.00hrs on 13.2.92 a long deceleration of the fetal heart was heard, and ARM was performed straight away.

Labour events

- 10.10 Cx 3cm dilated, thick. Station -2. ARM; very small amount of clear liquor seen.
- 10.35 Maternal oxygen given. Unsuccessful attempt at FBS.
- 11.00 Ranitidine 100mg maxolon 10mg iv given.
- 11.15 FBS; pH 7.34 BE -1.
- 11.40 Epidural begun.
- 12.45 VE Cx 5cm dilated, thick. Station -2,
- 13.15 Top up.
- 14.10 VE Cx 6cm dilated, thick. Station -2.
- 15.40 Top up.
- 16.15 VE Cx 7cm dilated, thick. Station -1.
- 16.30 Syntocinon started.
- 17.10 VE Cx 8cm dilated, oedematous. Station -1. Syntocinon stopped.
- 17.40 Top up.
- 18.10 VE Cx 8cm dilated. OP position. FBS; pH 7.27 BE -2
- 18.20 Syntocinon started.
- 18.50 Vomiting. Syntocinon stopped.
- 19.05 Vomiting.
- 19.10 VE Cx 8-9cm dilated. Station -1. Appearance of meconium.
- 19.25 Top up.
- 20.30 VE Cx 9cm dilated. Station 0. ROL position.
- 20.45 FBS; pH 7.12 BE -2
- 20.55 VE Cx fully dilated. Station; just below ischial spines.
Decision for delivery; trial of forceps. Top up. Maternal oxygen given.
- 21.10 Taken to theatre.
- 21.40 Male infant delivered by Kjelland forceps. Indication; fetal distress (poor CTG & low scalp pH).

Outcome

Birth weight 3.985 kg
Apgar 4 & 9
Cord gases pH 6.93 / 7.12 BD(ecf) 17 / 10



○ pH □ Cx Dilatation (cm)

Case 4 2025

Mrs S.P. is a 22 year old primigravida. She is a fit lady but smokes 5 cigarettes per day. Her pregnancy progressed well until 38 weeks, when mild hypertension without proteinuria was noted. At 40 weeks she developed 2+ proteinuria; BP 140/90-100. She was asymptomatic, reflexes were normal and both biochemistry and haematology were within normal limits. The fetus was judged to be of average size. Labour was induced on 17.9.91 on the grounds of proteinuric hypertension at term. Presentation cephalic 2/5. VE Cx 1cm dilated, part effaced. Prostin tablet 3mg was given at 21.10hrs. 6 hours later (03.00hrs) contractions were occurring regularly and Cx was now 1cm dilated, effaced and thin. Pethidine 100mg, phenergan 25mg was given im.

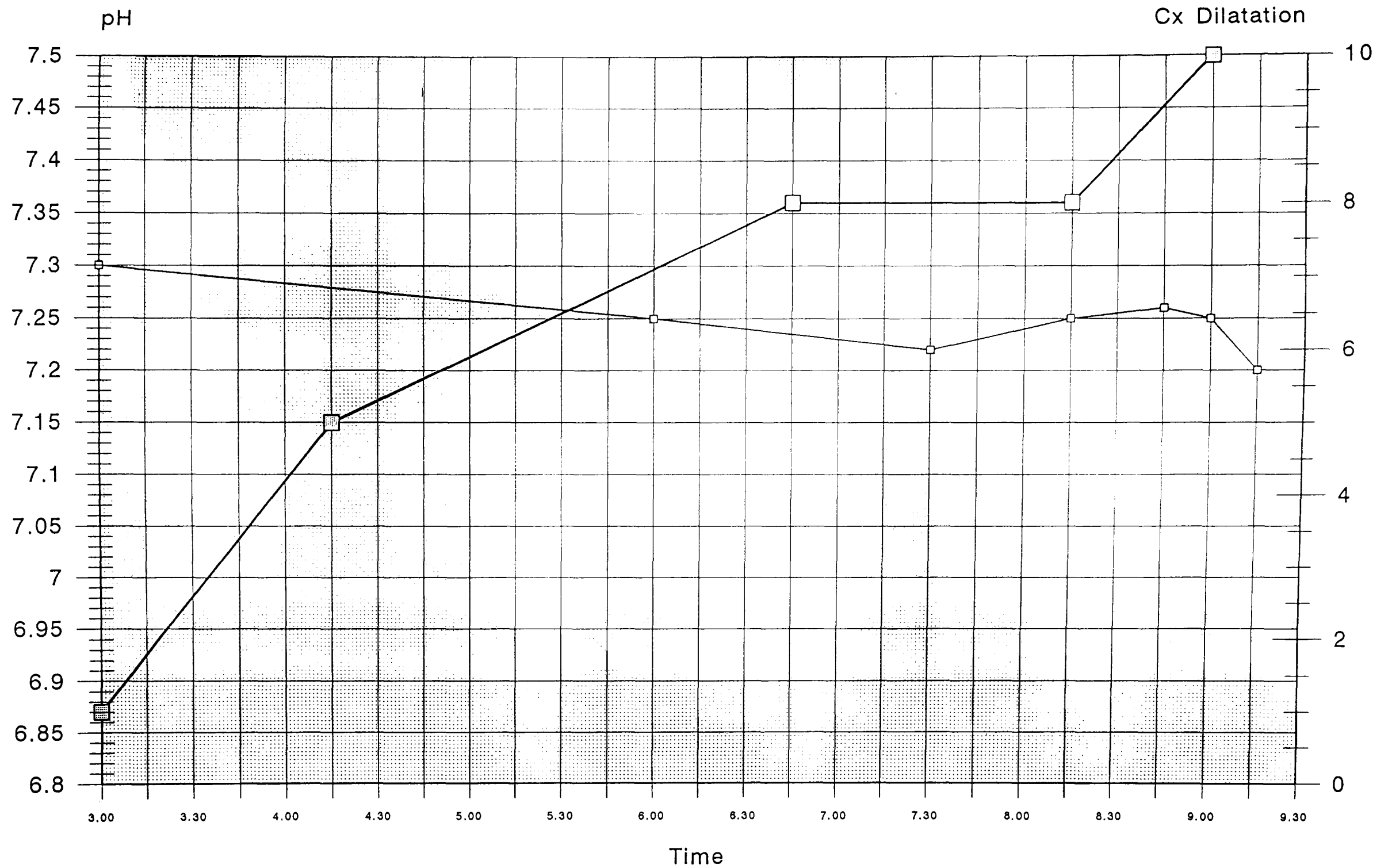
Labour events

Note; BP remained at 140-170/80-100 throughout labour.

- 04.15 VE Cx 5cm dilated, thin. Station -2. ARM; clear liquor. FSE applied.
- 04.45 Change of maternal position.
- 04.50 Epidural begun.
- 05.05 Epidural complete.
- 06.40 VE Cx 8cm dilated. Station -1. Direct OP position.
- 07.30 FBS; pH 7.22 BE -3
- 07.35 Bedpan.
- 07.45 Bedpan.
- 08.05 Bedpan.
- 08.20 VE Cx 8cm dilated.
- 08.40 FBS; pH 7.26 (no BE)
- 08.43 Top up.
- 08.55 Cephalic 0/5. VE Cx fully dilated. Station +1. OA position. Decision for forceps delivery.
- 09.17 Female infant delivered by Neville Barnes forceps. Indication; FH decelerations; borderline scalp pH.

Outcome

Birth weight 2.80kg
Apgar 5 & 9
Cord gases pH 7.20 / 7.29 BD(ecf) 5 / 4



○ pH □ Cx Dilatation (cm)

Case 5 1886

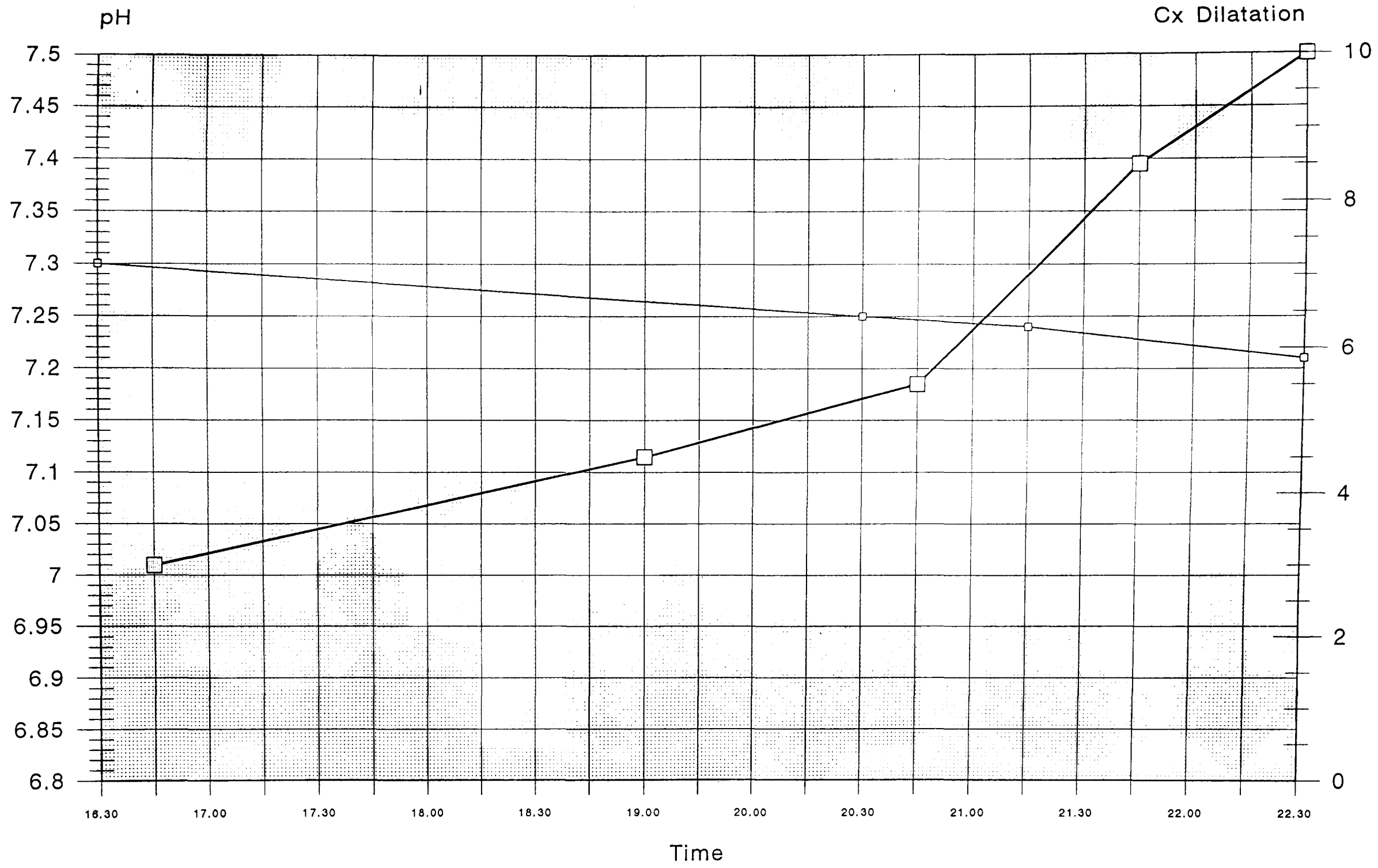
Mrs A.S. is a 30 year old lady expecting her second baby. She is fit and well, and does not smoke. Her first child was delivered by Caesarean section because of a footling breech presentation. (It was an emergency procedure, but prior to established labour.) The baby was full term, and weighed 8lb 4oz. Her current pregnancy progressed well, and a trial of vaginal delivery was planned. Labour was induced at 42 weeks on the grounds of postmaturity. On admission on 19.8.91, the baby was judged to be of an average size; presentation cephalic 2/5. On VE, Cx 3cm dilated, effaced. Station -1. ARM was performed; a small amount of clear liquor drained, and FSE applied. (16.50hrs)

Labour events

- 17.35 Pethidine 100mg, stemetil 12.5mg im.
- 18.35 Change of maternal position.
- 19.00 VE Cx 4-5cm dilated, thin. Station -1.
- 19.45 Epidural begun.
- 20.00 Epidural complete.
- 20.05 Syntocinon begun.
- 20.20 From now, for at least 1 hour, contractions very frequent (6-7:10)
- 20.50 VE Cx 5-6cm dilated, thin. Station -1.
- 21.20 FBS; pH 7.24 BE -2.
- 21.50 VE Cx 8-9cm dilated. Station 0.
- 21.55 Top up.
- 22.10 Mother catheterised.
- 22.30 VE Cx fully dilated.
- 22.35 Female infant delivered by Neville Barnes forceps. Indication; prolonged FH decerations. Thick fresh meconium noted at delivery.

Outcome

Birth weight 3.32kg
Apgar 5 & 9
Cord gases pH 7.21 / 7.30 BD(ecf) 4 / 3



○ pH □ Cx Dilatation (cm)

Case 6 1297

Miss J.H. is an 18 year old primigravida. She has no significant medical history but smokes 15 cigarettes per day. There were no antenatal problems and she was admitted in spontaneous labour at 40 weeks gestation, on 1.4.91. Presentation cephalic 3/5. On VE at 08.05hrs the cervix was 4cm dilated, effacing; station -1.

Labour events

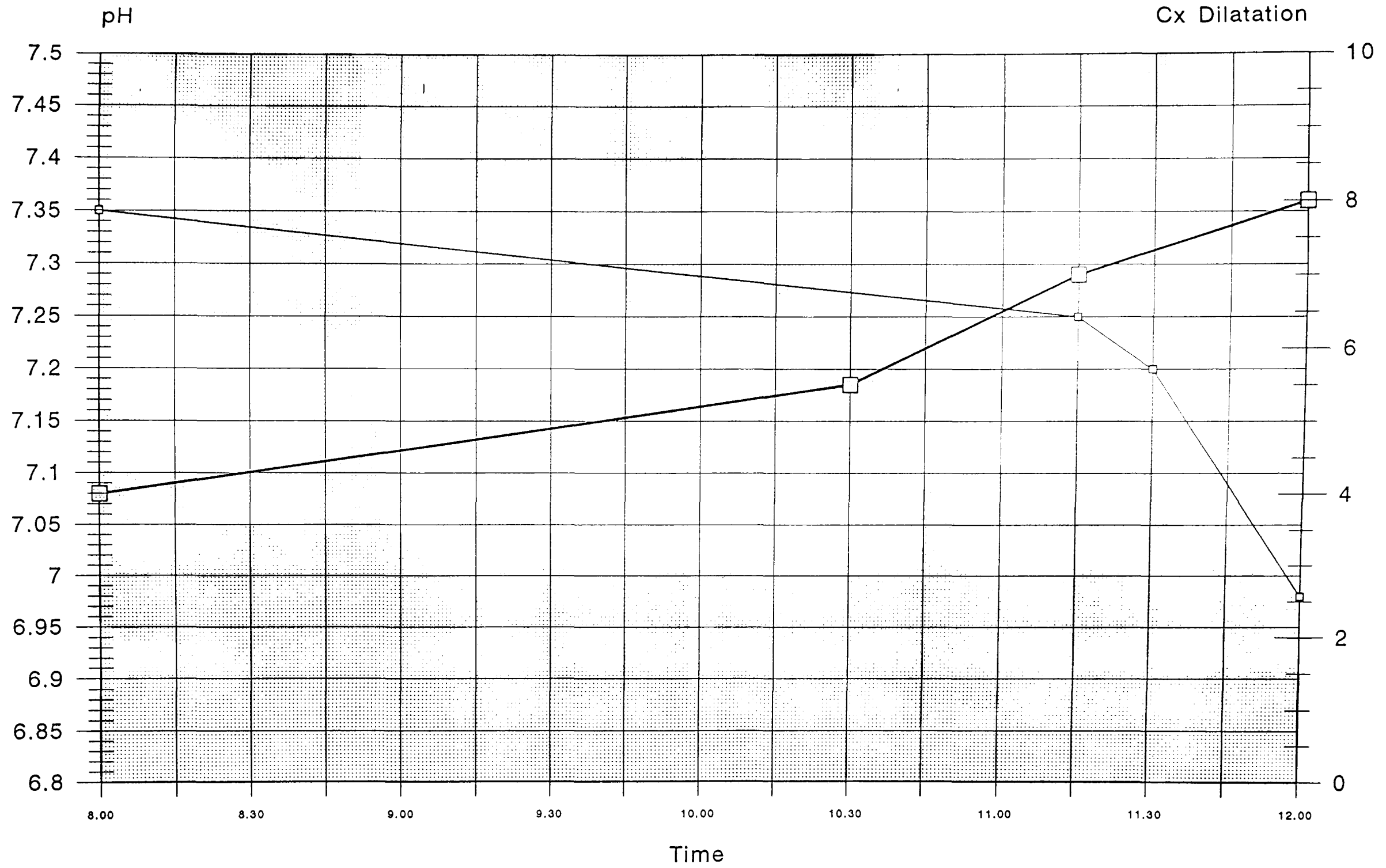
- 10.30 VE Cx 5-6cm dilated, thin. Station -1. LOT position. ARM; clear liquor. FSE applied.
- 11.10 Maternal position changed (left lateral) and facial oxygen given.
VE Cx 7cm dilated.
- 11.20 Decision for emergency Caesarean section.
- 11.52 Female infant delivered by C/S under GA. Indication; prolonged fetal bradycardia.
No cause for fetal distress identified.

Outcome

Birth weight 3.23kg

Apgar 4 & 9

Cord gases pH 6.96 / 7.00 BD(ecf) 14 / 10



□ pH □ Cx Dilatation (cm)

Case 7 393

Mrs J.B. is a 26 year old lady expecting her second baby. In 1987 she had a normal delivery of a baby boy, weighing 9lb 2oz. During the current pregnancy, there was clinical suspicion that the baby was large for dates. Investigation showed no evidence of diabetes. Labour was induced at 41 weeks for "postmaturity". When she was admitted for this purpose on 18.9.90 she had already started contracting at home, and the cervix was 3cm dilated. ARM was performed at 14.40hrs and clear liquor drained. She was monitored intermittently until 17.40hrs, when an FSE was applied. By this time the cervix was 4cm dilated; station -2.

Labour events

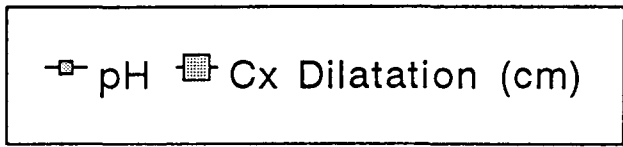
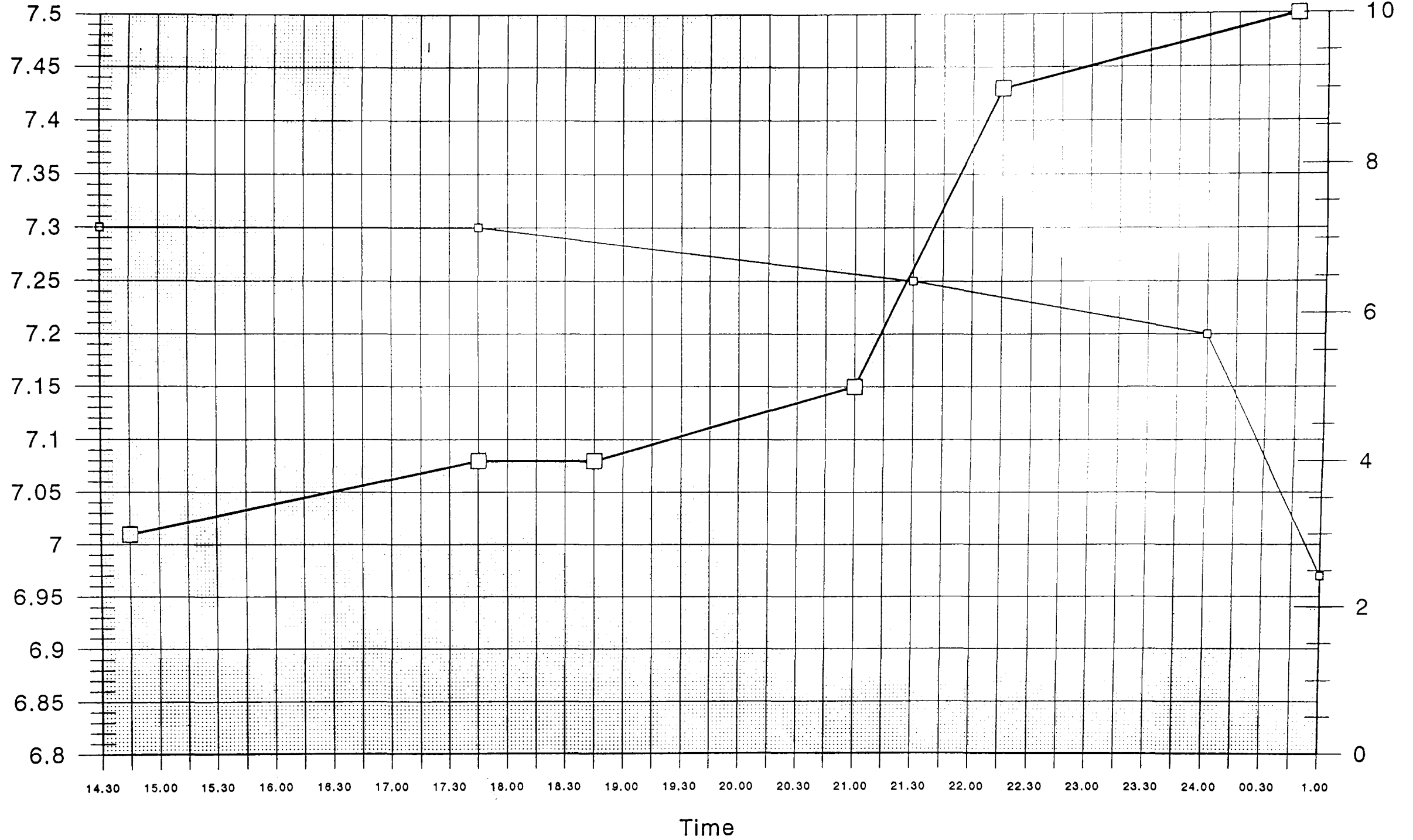
17.50 Pethidine 100mg, Stemetil 12.5mg im.
18.50 VE No change.
19.00 Epidural begun.
19.45 Epidural working.
20.40 Syntocinon started.
21.05 VE Cx 5cm dilated, thin. Station -2. OP position. FSE reapplied. (Lying flat for VE)
21.25 Hypotensive (BP 80/60) and nauseous. IV fluids increased.
22.00 Bedpan.
22.10 VE Cx 9cm dilated. Station -2. Direct OP position.
22.25 Top up.
00.45 VE Cx fully dilated. Station -1. OP position.
Pushing begun.
01.05 Normal delivery of male infant.

Outcome

Birth weight 3.85kg
Apgar 2 & 8
Cord gases pH 6.97 / 7.14 BD(ecf) 8 / 8

pH

Cx Dilatation



Case 8 713

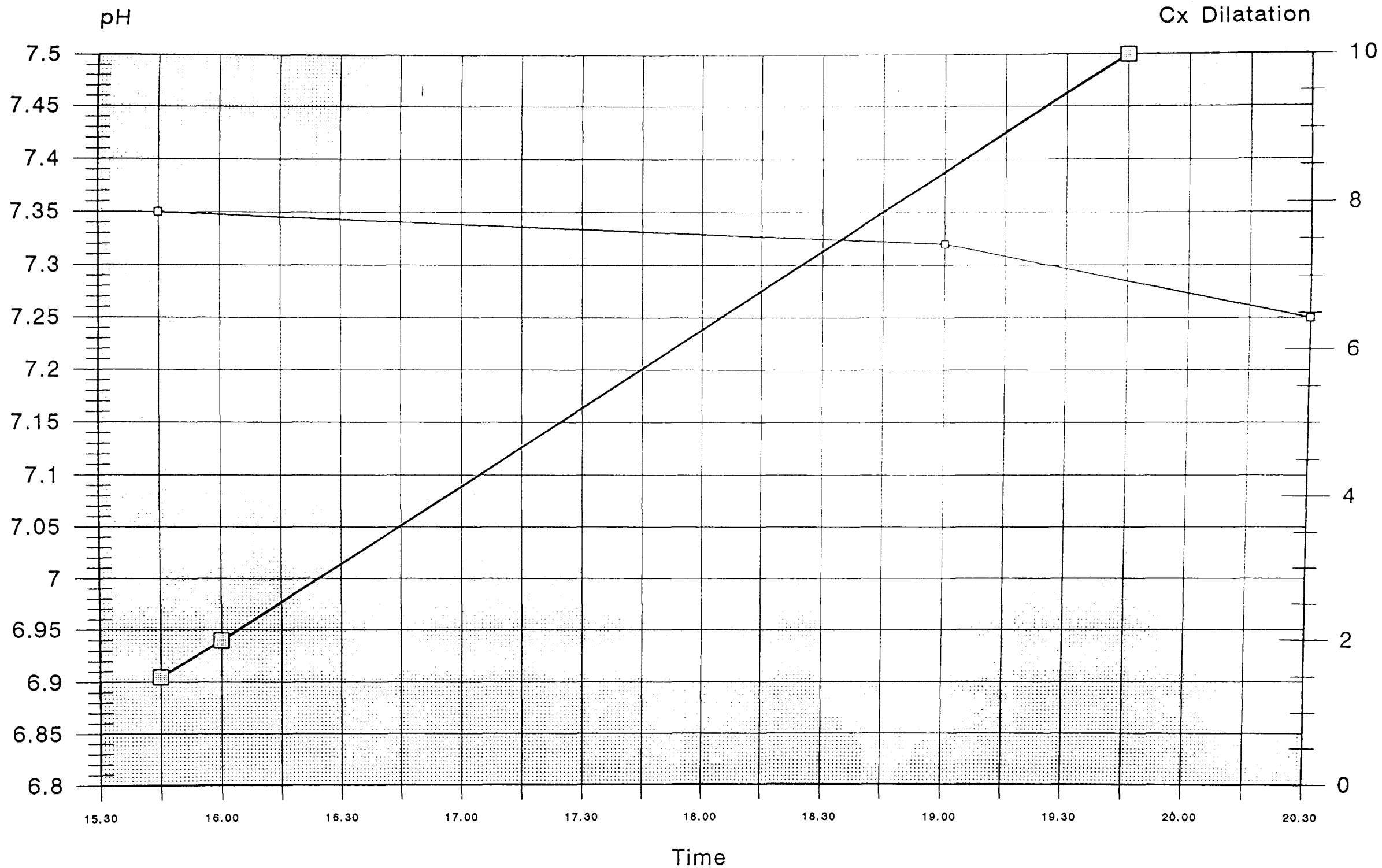
Mrs L.B. is a 27 year old primigravida; a fit nonsmoker. No antenatal problems were encountered, and she laboured spontaneously at 40 weeks. When she came to the labour ward, the membranes had ruptured 4 hours ago and regular contractions had begun. The uterus was term size, presentation cephalic 3/5. VE (12.30hrs on 26.11.91) Cx 1cm dilated, 2cm long. Station -3. Clear liquor noted. 3 hours later, at 15.40hrs, contractions were still mild and the cervix 1-2cm dilated, part effaced. FSE now applied, and syntocinon commenced.

Labour events

- 16.05 VE Cx 2cm dilated, effaced. Station -2.
- 16.23 Pethidine 100mg im.
- 16.50 Syntocinon begun.
- 17.00 Maternal position changed, oxygen given. Syntocinon stopped.
- 17.30 Syntocinon restarted.
- 18.40 Epidural begun.
- 19.00 Maternal hypotension. IV fluids increased.
- 19.20 Epidural complete.
- 19.40 PV bleeding; small clots. VE Cx fully dilated. OA position. Station +1
Decision for forceps.
- 20.00 Top up.
- 20.32 Male infant delivered by Neville Barnes forceps. Indication; PV bleeding at full dilatation.

Outcome

Birth weight 3.56kg
Apgar 8 & 9
Cord gases pH 7.25 / 7.31 BD(ecf) 4 / 6



□ pH ■ Cx Dilatation (cm)

Case 9 1129

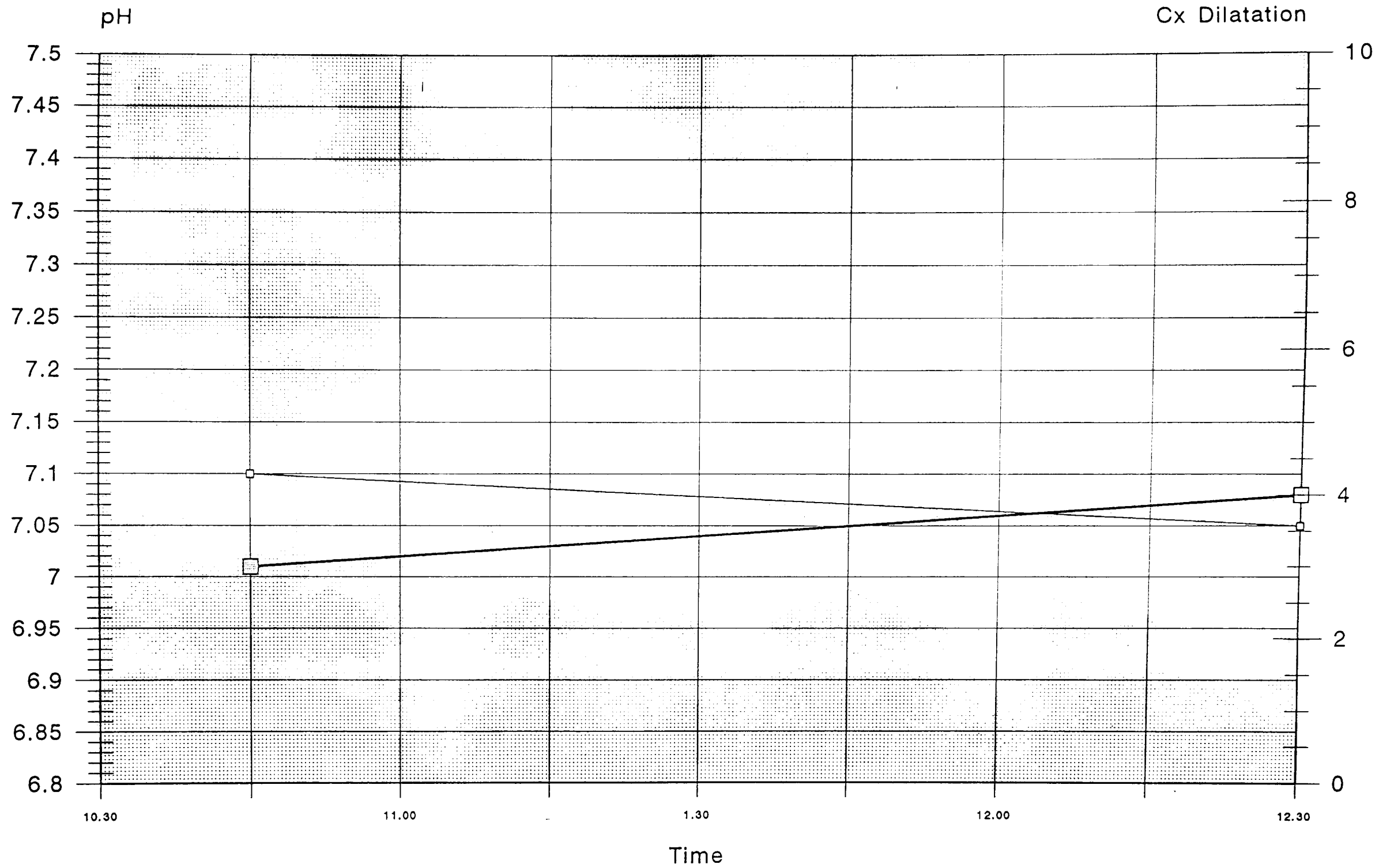
Miss S.W. is a 28 year old primigravida. She is a fit lady but smokes 10 cigarettes per day. There were no antenatal problems, and she laboured spontaneously at 39 weeks gestation. She was admitted on 27/2/91 at 10.50hrs in early labour Uterus term size. Cephalic 3/5. VE Cx 3cm dilated, 0.5cm thick. Station -3.

Labour events

- 11.35 VE as described. ARM; thick meconium. Unsuccessful attempt at FBS.
Maternal oxygen given.
Decision for C/S.
- 12.24 Female infant delivered by emergency C/S under GA. Indication; poor CTG very early in labour. Scalp pH not possible.

Outcome

Birth weight 2.38kg
Apgar 5 & 7
Cord gases pH 7.03 / 7.10 BD(ecf) 13 / 13



Case 10 518

Mrs V.C. is a 33 year old primigravida; a fit lady and a nonsmoker. Her pregnancy was straightforward, and she was admitted with spontaneous rupture of the membranes at 42 weeks, on 23.10.90. On admission the fetus was felt to be well grown, cephalic presentation 2/5. VE (at 10.30hrs) Cx 1cm dilated, uneffaced. Station -2. Clear liquor seen. Labour did not establish over the next few hours, and prostin gel 2mg was given at 14.50hrs. Intermittent monitoring was performed. By 19.35hrs contractions were becoming regular; the cervix now 4cm dilated, fully effaced. Station -2. Forewater rupture performed; liquor stained with fresh meconium.

Labour events

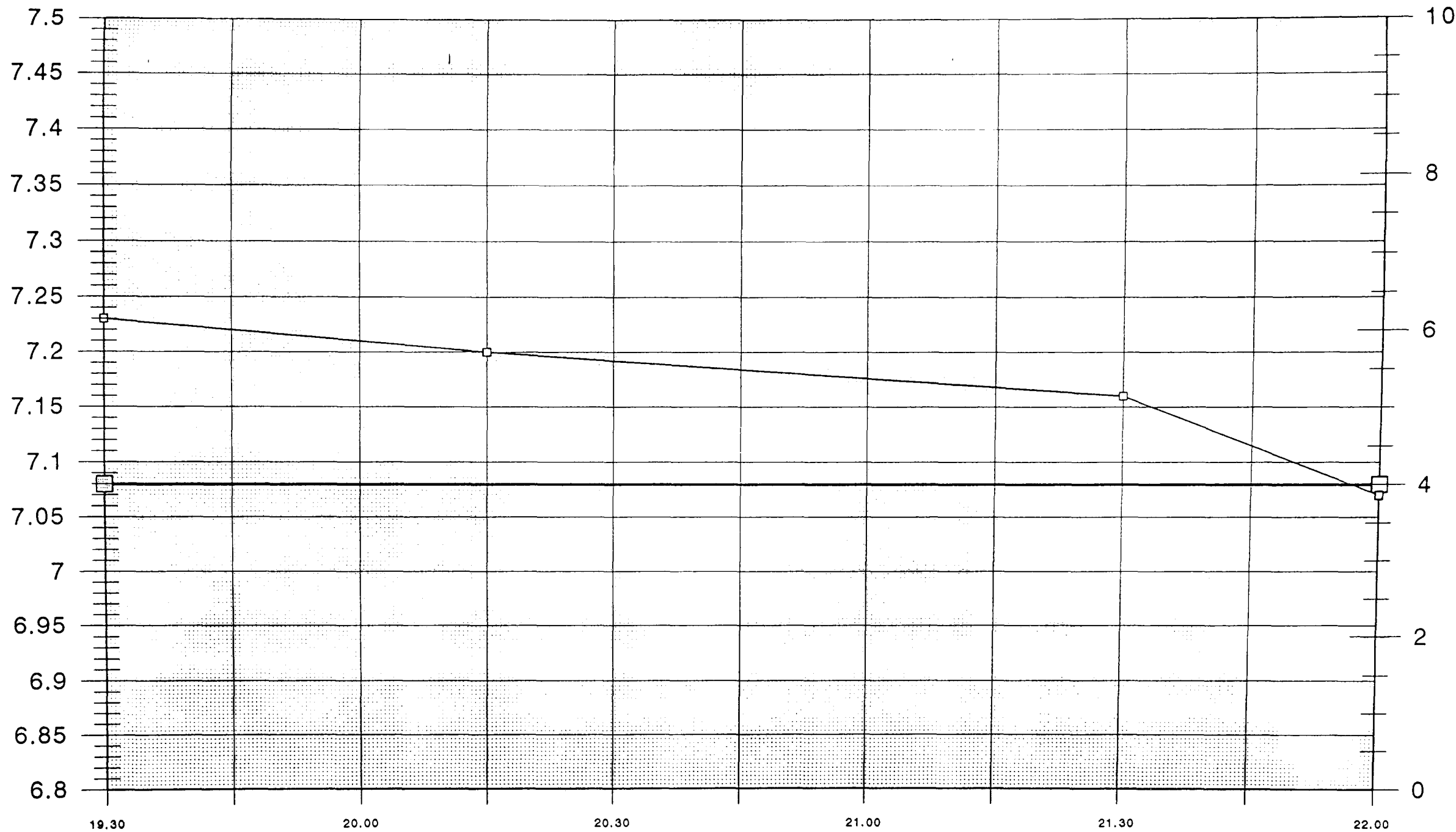
- 20.00 Pethidine 100mg, phenergan 25mg im.
- 20.20 Bedpan.
- 21.00 Maternal position changed, oxygen given.
- 21.24 VE Cx 4cm dilated. FBS; pH 7.16 BE -10
Decision for C/S
- 22.00 Female infant delivered by emergency C/S under GA. Indication; low scalp pH at 4cm dilatation.

Outcome

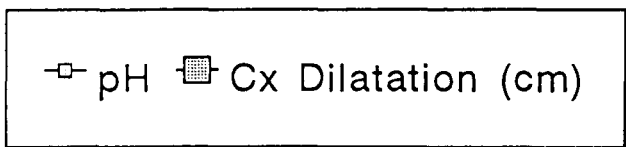
Birth weight 3.83kg
Apgar 3 & 7 (8 at 15 minutes)
Cords pH 7.05 / 7.12 BD(ecf) 5 / 6

pH

Cx Dilatation



Time



Case 11 1526

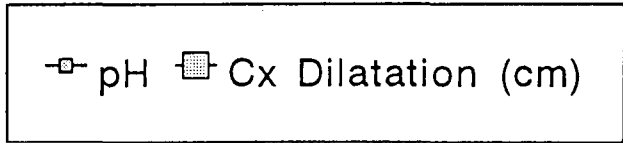
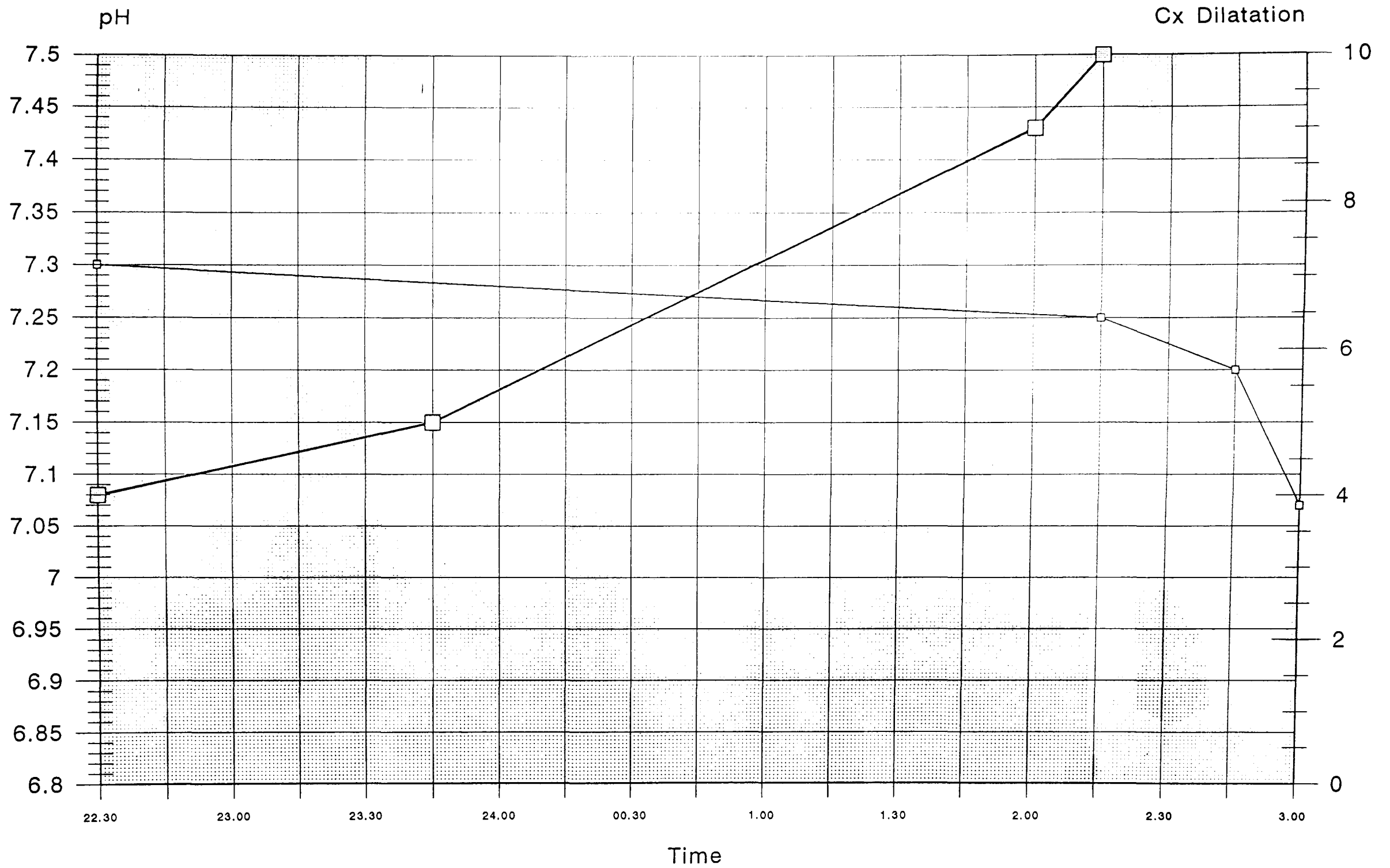
Mrs C.L. is a 20 year old primigravida. She is a fit lady but smokes 10 cigarettes per day. There were no antenatal problems, and she laboured spontaneously at 37 weeks. On admission (22.5.91) the uterus was appropriate for dates, presentation cephalic 2/5. VE (at 22.30hrs) Cx 4cm dilated, effaced and thin. Station 0. Clear liquor already draining.

Labour events

23.50 VE Cx 5cm dilated. Station 0. FSE applied.
00.15 Epidural begun.
00.35 Epidural complete.
02.00 VE Cx 9cm dilated. Station 0.
02.15 VE Cx fully dilated. Pushing begun.
03.05 Normal delivery of male infant.

Outcome

Birth weight 3.11kg
Apgar 9 & 9
Cord gases pH 7.15 BD(ecf) - / 6



Case 12 1316

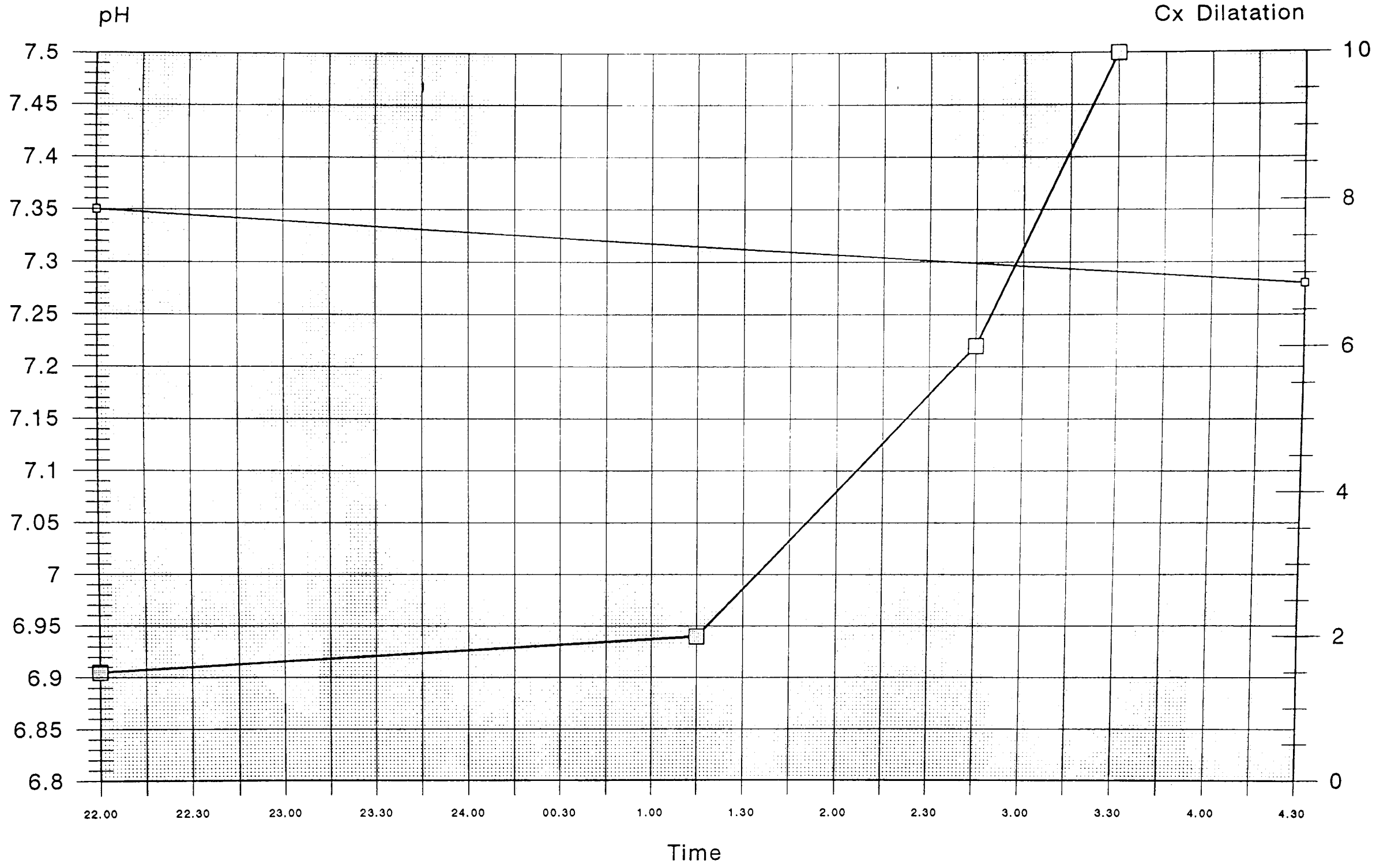
Miss S.W. is a 18 year old primigravida. She suffers from Crohn's disease which is well controlled on mesalazine. She does not smoke. There were no antenatal problems. Labour was induced at 40 weeks on the grounds of a severe pregnancy-related rash. The baby was felt to be adequately grown, presentation cephalic 2/5. The cervix was closed and uneffaced. Prostin gel 1mg was given on 6.4.91 at 11.30hrs, and a further 1mg at 18.15hrs. By 22.00hrs contractions were occurring regularly and the cervix was by now 1-2cm dilated, partly effaced.

Labour events

- 01.10 VE Cx 2cm dilated, effacing. ARM; clear liquor. FSE applied.
- 01.20 Pethidine 100mg, phenergan 25mg im.
- 02.50 VE Cx 6cm dilated, thin. Station +1. FSE reapplied.
- 03.30 VE Cx fully dilated. Pushing begun.
- 04.30 Normal delivery of female infant.

Outcome

Birth weight 3.66kg
Apgar 9 & 9
Cord gases pH 7.28 / 7.39 BD(ecf) 3 / 3



□ pH □ Cx Dilatation (cm)

Case 13 2315

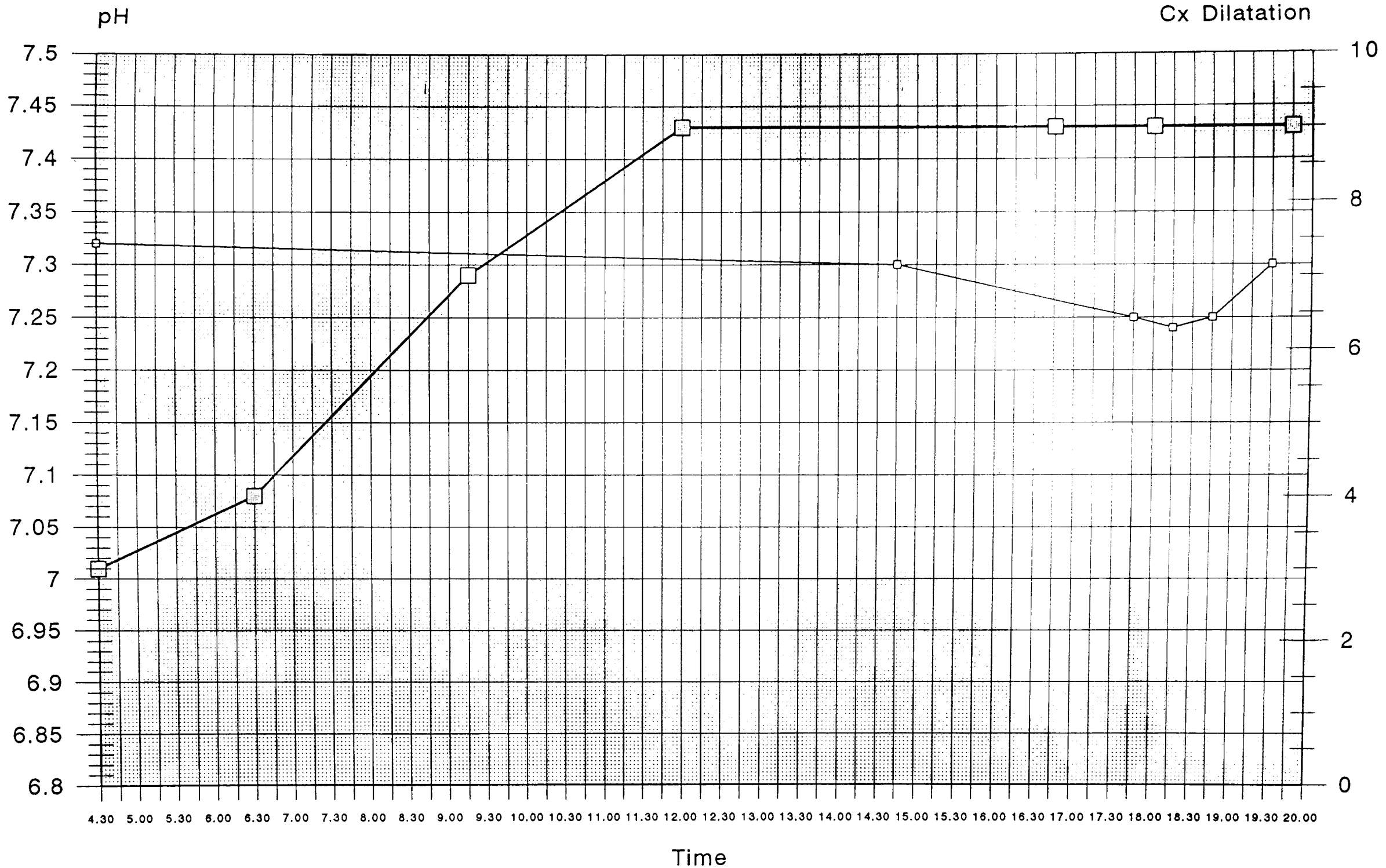
Mrs C.H. is a 37 year old primigravida; a fit lady, and a nonsmoker. Biochemical testing suggested a low risk of Down's syndrome; amniocentesis was declined. No problems were encountered during the pregnancy, and she laboured spontaneously at 39 weeks. She was admitted on 18.11.91 at 16.45hrs. The membranes had ruptured 8 hours ago, and mild contractions were just starting. The uterus was term size, presentation cephalic 3/5. VE Cx 1cm dilated, with plenty of clear liquor noted. She was monitored intermittently until labour established, returning to the labour ward at 02.15hrs on 19.11.91.

Labour events

- 04.30 VE Cx 3cm dilated, effaced and thin. Station -1.
Pethidine 100mg, phenergan 25mg given im.
- 06.30 VE Cx 4cm dilated. Station -1. FSE applied.
- 07.15 Epidural inserted.
- 07.50 Syntocinon begun.
- 09.10 VE Cx 7cm dilated. Station 0.
- 09.50 Contractions too frequent; syntocinon reduced.
- 12.00 VE Cx 9cm dilated. Station 0.
- 14.50 VE Cx thought to be fully dilated. Station 0. LOA position. Pushing begun.
- 16.15 No progress. Top up.
- 16.50 VE Cx 9cm dilated. Station 0. Caput + moulding ++. LOA position.
- 18.00 VE no change.
- 18.19 FBS; pH 7.24 BE -8. Decision for C/S.
- 18.45 Top up.
- 19.00 Syntocinon discontinued.
- 19.45 Female infant delivered by emergency C/S under epidural. Indication; failure to progress in labour despite augmentation. Overdistended lower segment noted at operation, and liquor stained with meconium.

Outcome

Birth weight 3.62kg
Apgar 5 & 9
Cord gases pH 7.28 / 7.32 BD(ecf) 3 / 4



□ pH ○ Cx Dilatation (cm)

Case 14 1037

Mrs S.B. is a 29 year old lady expecting her second baby. She is fit and well, and a nonsmoker. Her first child was born normally at term, and weighed 7lb 10oz. There were no antenatal problems but she was admitted for induction of labour at 43 weeks. The baby was felt to be well grown, presentation cephalic 2/5. VE Cx 3cm dilated. Admission CTG started at 15.50hrs.

Labour events

5.2.91

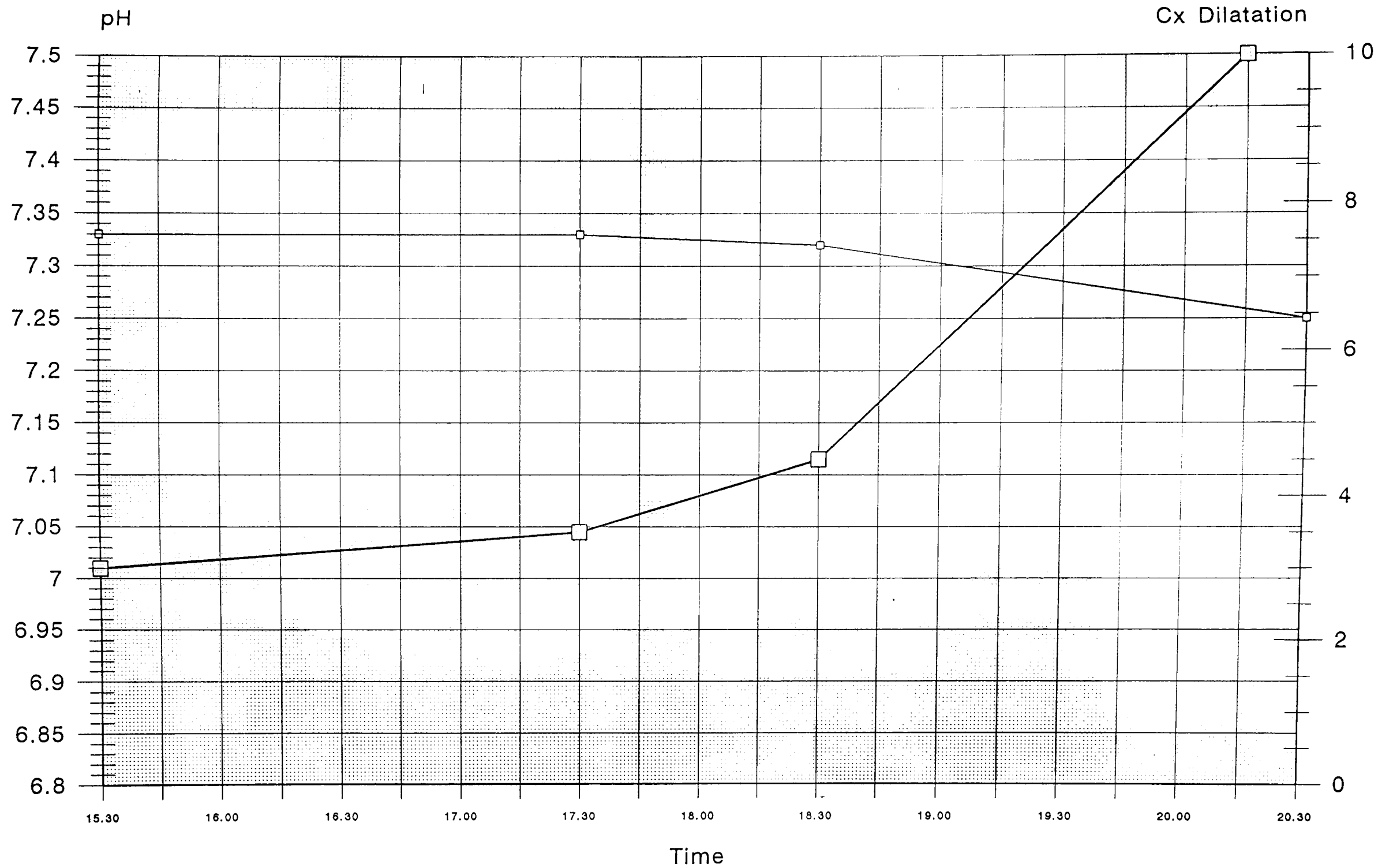
- 17.30 VE Cx 3-4cm dilated, 0.5cm thick. Station -1. ARM; minimal clear liquor.
- 18.30 VE Cx 4-5cm dilated. FBS; pH 7.32 BE 0
- 19.05 Syntocinon begun.
- 19.20 Stemetil 12.5mg im.
- 19.30 Vomiting.
- 19.35 Bedpan.
- 20.05 Syntocinon stopped.
- 20.08 VE Cx fully dilated; vertex visible.
- 20.34 Normal delivery of female infant.

Outcome

Birth weight 2.94kg

Apgar 5 & 9

Cord gases pH 7.25 / 7.32 BD(ecf) 0 / 1



□ pH □ Cx Dilatation (cm)

Case 15²⁵¹

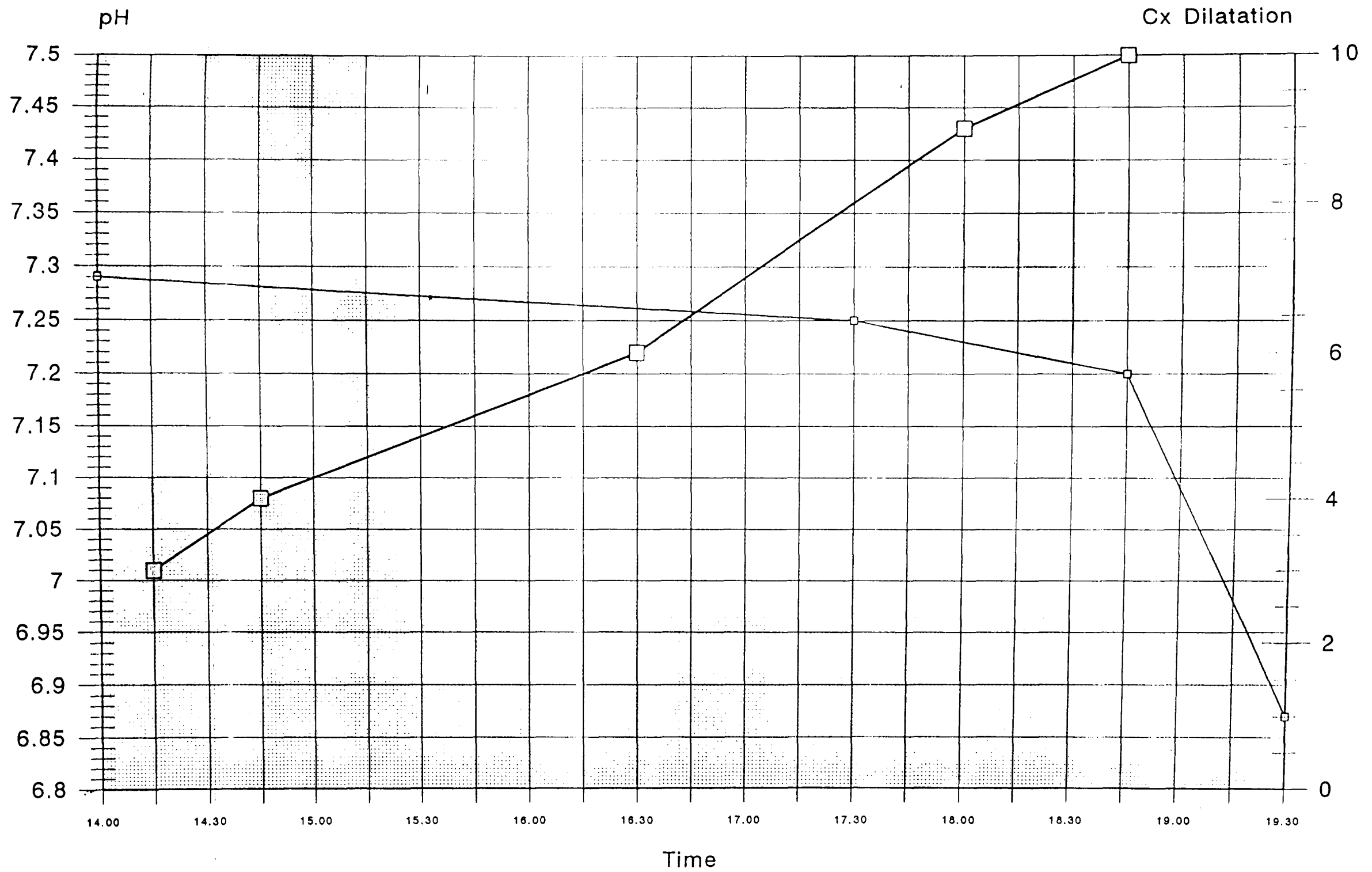
Mrs C.W. is a 27 year old primigravida; fit, and a nonsmoker. Her antenatal progress was good and she was admitted in spontaneous labour at 41 weeks gestation. At this stage (08.00hrs on 21.8.90) contractions were mild and the cervix closed. She was monitored intermittently until 14.00hrs, when labour was established.

Labour events

- 14.10 Now contracting more regularly. Cephalic 2/5. VE Cx 3cm dilated, 1cm thick. Station -1. ARM performed; thick meconium drained. FSE applied.
- 14.40 FSE fell off; reapplied 5 minutes later. Cx now 4cm dilated. Station -1.
- 16.30 VE Cx 6cm dilated & thin. Station -1. LOA position.
- 18.00 VE Cx 9cm dilated. Station 0. LOA position.
- 18.45 VE Cx fully dilated. Pushing begun.
- 19.30 Normal delivery of female infant.

Outcome

Birth weight 3.28kg
Apgar 5 & 7 (8at 10 minutes)
Cord gases pH 6.87 / 7.08 BD(ecf) 16 / 3



□ pH □ Cx Dilatation (cm)

Agreement results matrix for case 15

		REVIEWER																																					
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	I1	I2	J1	J2	K1	K2	L1	L2	M1	M2	N1	N2	O1	O2	P1	P2	Q1	Q2	S1	S2	R1	R2
REVIEWER	A1	100	95	10	21	76	64	91	95	96	97	70	61	95	95	2	33	81	82	70	72	63	92	61	96	97	97	93	70	72	94	68	80	93	93	88	88	68	80
	A2	95	100	9	19	81	68	96	98	95	92	65	58	96	89	2	32	76	79	65	70	60	97	56	99	91	91	88	65	73	98	67	75	88	88	93	93	69	79
	B1	10	9	100	75	26	36	8	10	10	10	41	43	10	11	39	64	27	27	39	38	44	9	24	9	10	10	11	37	40	9	38	17	11	11	8	8	33	23
	B2	21	19	75	100	36	45	17	20	22	22	47	54	21	23	39	51	42	42	46	52	53	18	51	19	22	22	23	47	49	19	49	25	23	23	17	17	38	33
	C1	76	81	26	36	100	85	79	79	79	76	74	72	78	76	31	44	91	94	79	74	66	80	51	79	73	73	76	81	78	80	85	75	76	76	77	77	80	93
	C2	64	68	36	45	85	100	71	66	64	61	73	79	63	61	29	43	80	83	82	73	70	69	42	68	59	59	61	88	73	69	74	62	61	61	70	70	69	82
	D1	91	96	8	17	79	71	100	93	87	84	59	53	89	81	1	29	70	73	59	68	58	99	50	97	83	83	79	59	68	97	65	67	79	79	99	99	68	77
	D2	95	98	10	20	79	66	93	100	95	94	66	59	97	91	2	32	77	79	67	73	59	95	58	97	93	93	89	67	72	98	66	76	89	89	91	91	69	78
	E1	96	95	10	22	79	64	87	95	100	98	71	64	98	97	3	34	81	81	71	69	63	90	62	92	97	97	95	71	73	93	70	82	95	95	85	85	71	79
	E2	97	92	10	22	76	61	84	94	98	100	72	63	97	99	3	34	82	82	72	71	62	86	63	91	99	99	97	72	72	91	68	83	97	97	82	82	68	78
	F1	70	65	41	47	74	73	59	66	71	72	100	78	69	74	17	48	86	84	96	87	91	61	65	65	72	72	75	87	93	64	84	66	75	75	57	57	61	69
	F2	61	58	43	54	72	79	53	59	64	63	78	100	61	65	30	59	72	70	85	76	84	55	67	57	63	63	65	88	76	57	87	72	65	65	51	51	65	76
	G1	95	96	10	21	78	63	89	97	98	97	69	61	100	94	2	33	78	80	69	70	61	91	60	96	95	95	92	69	72	95	66	79	92	92	87	87	68	76
	G2	95	89	11	23	76	61	81	91	97	99	74	65	94	100	4	34	84	82	74	68	62	83	64	88	98	98	99	74	70	88	70	85	99	99	79	79	70	79
	H1	2	2	39	39	31	29	1	2	3	3	17	30	2	4	100	24	25	25	24	15	14	2	4	2	2	4	29	16	2	28	29	4	4	0	0	54	33	
	H2	33	32	64	51	44	43	29	32	34	34	48	59	33	34	24	100	41	42	51	46	52	31	52	31	33	33	35	51	48	32	56	42	35	35	28	28	39	45
	I1	81	76	27	42	91	80	70	77	81	82	86	72	78	84	25	41	100	98	88	81	72	71	57	75	82	82	85	86	82	75	84	74	85	85	67	67	74	88
	I2	82	79	27	42	94	83	73	79	81	82	84	70	80	82	25	42	98	100	86	82	72	75	56	79	80	80	82	84	84	78	83	72	82	82	71	71	73	88
	J1	70	65	39	46	79	82	59	67	71	72	96	85	69	74	24	51	88	86	100	78	87	61	58	65	72	72	75	96	87	64	81	66	75	75	57	57	66	77
	J2	72	70	38	52	74	73	68	73	69	71	87	76	70	68	15	46	81	82	78	100	88	68	69	71	70	70	67	70	95	71	86	59	67	67	67	67	58	69
	K1	63	60	44	53	66	70	58	59	63	62	91	84	61	62	14	52	72	72	87	88	100	59	66	60	63	63	61	78	92	59	84	54	61	61	56	56	54	59
	K2	92	97	9	18	80	69	99	95	90	86	61	55	91	83	2	31	71	75	61	68	59	100	52	98	86	86	82	61	69	98	64	69	82	82	97	97	69	79
	L1	61	56	24	51	51	42	50	58	62	63	65	67	60	64	4	52	57	56	58	69	66	52	100	56	63	63	65	52	66	55	71	58	65	65	49	49	49	54
	L2	96	99	9	19	79	68	97	97	92	91	65	57	96	88	2	31	75	79	65	71	60	98	56	100	91	91	87	65	71	98	66	74	87	87	95	95	68	79
	M1	97	91	10	22	73	59	83	93	97	99	72	63	95	98	2	33	82	80	72	70	63	86	63	91	100	100	97	72	70	89	68	83	97	97	81	81	68	77
	M2	97	91	10	22	73	59	83	93	97	99	72	63	95	98	2	33	82	80	72	70	63	86	63	91	100	100	97	72	70	89	68	83	97	97	81	81	68	77
	N1	93	88	11	23	76	61	79	89	95	97	75	65	92	99	4	35	85	82	75	67	61	82	65	87	97	97	100	75	69	86	71	86	100	100	77	77	70	80
	N2	70	65	37	47	81	88	59	67	71	72	87	88	69	74	29	51	86	84	96	70	78	61	52	65	72	72	75	100	75	64	76	73	75	75	56	56	69	82
	O1	72	73	40	49	78	73	68	72	73	72	93	76	72	70	16	48	82	84	87	95	92	69	66	71	70	70	69	75	100	72	88	61	69	69	66	66	60	69
	O2	94	98	9	19	80	69	97	98	93	91	64	57	95	88	2	32	75	78	64	71	59	98	55	98	89	89	86	64	72	100	66	73	86	86	95	95	70	79
	P1	68	67	38	49	85	74	65	66	70	68	84	87	66	70	28	56	84	83	81	86	84	64	71	66	68	68	71	76	88	66	100	70	71	71	64	64	71	82
	P2	80	75	17	25	75	62	67	76	82	83	66	72	79	85	29	42	74	72	66	59	54	69	58	74	83	83	86	73	61	73	70	100	86	86	65	65	74	86
Q1	93	88	11	23	76	61	79	89	95	97	75	65	92	99	4	35	85	82	75	67	61	82	65	87	97	97	100	75	69	86	71	86	100	100	77	77	70	80	
Q2	93	88	11	23	76	61	79	89	95	97	75	65	92	99	4	35	85	82	75	67	61	82	65	87	97	97	100	75	69	86	71	86	100	100	77	77	70	80	
S1	88	93	8	17	77	70	99	91	85	82	57	51	87	79	0	28	67	71	57	67	56	97	49	95	81	81	77	56	66	95	64	65	77	77	100	100	67	76	
S2	88	93	8	17	77	70	99	91	85	82	57	51	87	79	0	28	67	71	57	67	56	97	49	95	81	81	77	56	66	95	64	65	77	77	100	100	67	76	
R1	68	69	33	38	80	69	68	69	71	68	61	65	68	70	54	39	74	73	66	58	54	69	49	68	68	68	70	69	60	70	71	74	70	70	67	67	100	80	
R2	80	79	23	33	93	82	77	78	79	78	69	76	76	79	33	45	88	88	77	69	59	79	54	79	77	77	80	82	69	79	82	86	80	80	76	76	80	100	

Case 16 1459

Miss S.H. is a 28 year old primigravida; a fit lady and a nonsmoker. Her antenatal progress was good. She was noted to be of short stature (4ft 11in); clinically her pelvis was thought to have a small outlet but Xray pelvimetry demonstrated good AP diameters. The fetus was felt to be of an average size; cephalic, 4-5/5 palpable per abdomen. Labour was induced at 42 weeks on the grounds of postmaturity. On admission (4.5.91 at 17.10hrs) the cervix was 1cm dilated and soft; station -3, and Prostin gel 2mg was given PV.

Labour events

Note; mild maternal pyrexia (37.5 to 37.7°C) from 05.30 until delivery.

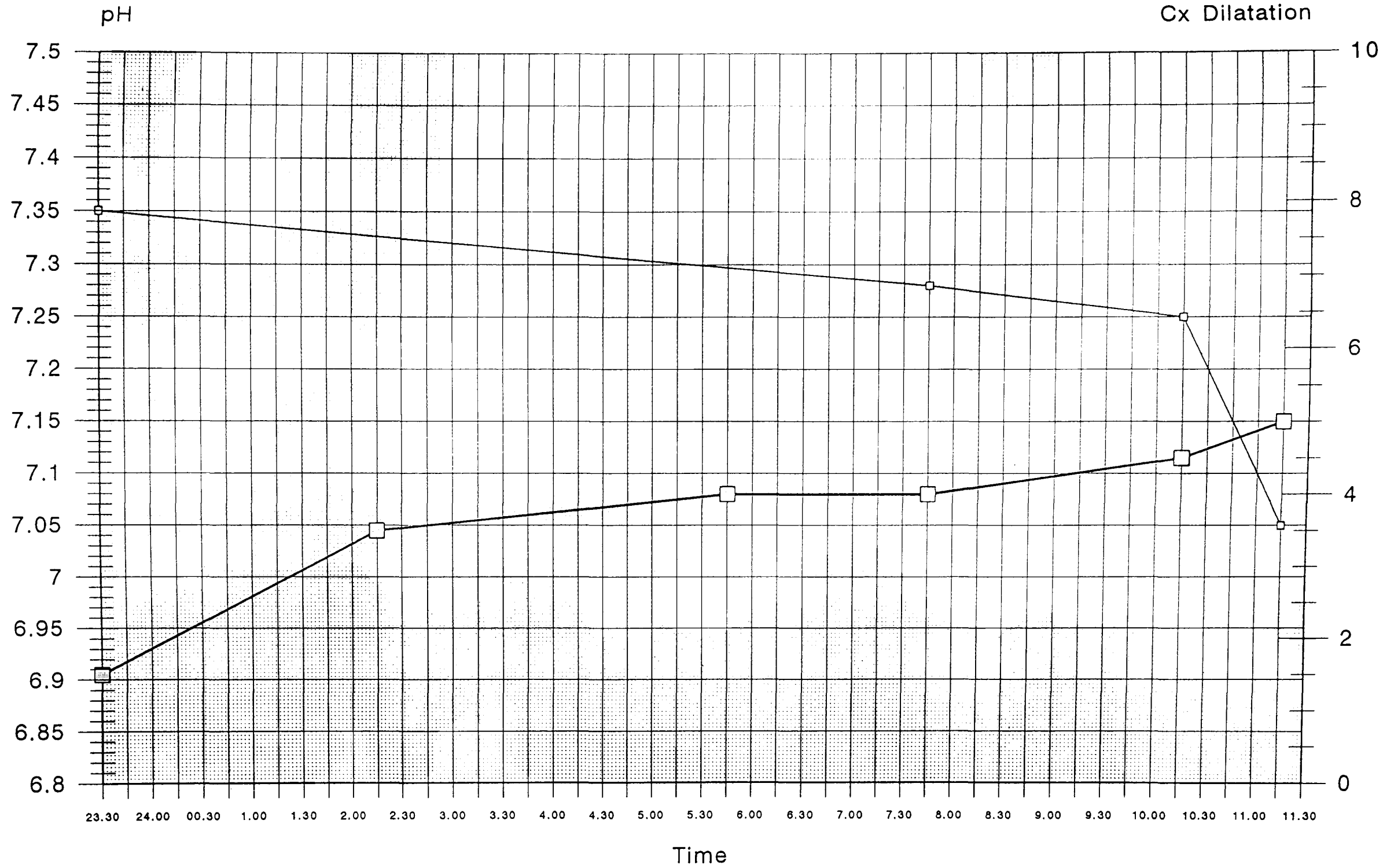
- 23.30 Contracting 4:10. VE Cx 1.5cm dilated, effaced, 0.5cm thick. Station -2 / -3.
- 00.20 Pethidine 100mg, stemetil 12.5mg im.
- 02.10 VE Cx 3-4cm dilated, thin. Station -2. ARM; clear liquor. FSE applied.
- 03.15 Epidural begun.
- 03.45 Epidural working.
- 05.40 VE Cx 4cm dilated, thin but poorly applied to presenting part. Station -2.
- 06.10 Top up.
- 06.30 Syntocinon started.
- 07.45 VE Cx 4cm dilated. Station -2. IUPD inserted.
FBS; pH 7.28 BE -8
- 08.30 Top up.
- 09.20 Appearance of thick fresh meconium.
- 10.15 VE Cx 4-5cm dilated, still quite thick.
FBS; pH 7.25 BE -13 Decision for Caesarean section.
- 10.20 Syntocinon discontinued.
- 10.30 Top up.
- 10.40 Taken to theatre.
- 11.17 Male infant delivered by C/S under epidural. Indication; falling fetal scalp pH and poor progress in labour.

Outcome

Birth weight 3.50kg

Apgar 7 & 9

Cord gases pH 7.05 / 7.13 BD(ecf) - / 7



○ pH □ Cx Dilatation (cm)

Agreement results matrix for case 16

		REVIEWER																																					
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	I1	I2	J1	J2	K1	K2	L1	L2	M1	M2	N1	N2	O1	O2	P1	P2	Q1	Q2	S1	S2	R1	R2
REVIEWER	A1	100	66	18	20	45	54	58	31	60	41	35	63	77	68	74	79	49	60	38	42	42	42	53	75	50	58	39	50	77	68	50	45	11	13	52	52	16	62
	A2	66	100	20	26	41	84	64	24	71	39	40	77	55	45	77	81	54	58	39	49	45	39	52	53	52	56	36	48	56	61	72	73	11	13	60	60	22	51
	B1	18	20	100	57	33	27	22	22	22	33	22	24	10	7	23	27	21	19	40	14	22	28	13	12	13	22	15	22	19	15	24	19	73	65	34	34	75	23
	B2	20	26	57	100	35	35	23	38	28	53	37	33	15	20	32	32	31	25	45	19	33	34	23	23	24	29	33	28	28	23	28	29	52	60	43	43	32	26
	C1	45	41	33	35	100	48	56	52	52	68	36	42	37	38	39	44	41	36	71	56	77	57	33	37	34	42	46	80	43	36	42	35	24	23	59	59	21	42
	C2	54	84	27	35	48	100	66	33	71	43	31	77	50	42	77	74	55	47	43	44	41	34	46	47	48	49	27	36	52	49	73	68	20	19	58	58	18	57
	D1	58	64	22	23	56	66	100	36	93	44	26	79	67	68	55	55	42	42	39	65	43	35	40	41	32	46	30	44	53	41	64	55	17	17	44	44	13	61
	D2	31	24	22	38	52	33	36	100	33	80	61	30	27	43	30	35	30	25	57	25	72	75	21	23	20	35	79	54	31	24	22	19	40	44	40	40	9	20
	E1	60	71	22	28	52	71	93	33	100	44	24	84	68	68	72	66	50	47	41	58	46	38	47	45	42	50	28	44	58	47	61	54	20	21	50	50	14	60
	E2	41	39	33	53	68	43	44	80	44	100	73	43	34	44	40	48	40	40	73	38	78	78	34	37	33	45	82	72	37	35	35	35	28	32	59	59	17	27
	F1	35	40	22	37	36	31	26	61	24	73	100	30	25	38	27	38	28	37	42	34	56	79	28	31	28	36	84	50	32	35	22	34	12	17	49	49	20	16
	F2	63	77	24	33	42	77	79	30	84	43	30	100	62	55	72	71	53	51	36	43	41	38	50	51	42	50	27	37	47	45	75	73	22	22	54	54	17	55
	G1	77	55	10	15	37	50	67	27	68	34	25	62	100	85	59	56	49	54	29	48	38	30	52	56	51	56	30	38	68	58	46	44	4	7	38	38	10	71
	G2	68	45	7	20	38	42	68	43	68	44	38	55	85	100	47	48	52	59	26	49	38	34	58	53	52	59	45	39	68	61	34	33	5	9	33	33	8	58
	H1	74	77	23	32	39	77	55	30	72	40	27	72	59	47	100	90	61	59	45	45	47	39	58	66	56	61	28	42	71	57	53	49	18	21	60	59	18	54
	H2	79	81	27	32	44	74	55	35	66	48	38	71	56	48	90	100	57	64	45	45	46	44	57	67	53	66	38	46	69	60	47	48	17	21	59	59	19	54
	I1	49	54	21	31	41	55	42	30	50	40	28	53	49	52	61	57	100	90	43	40	46	37	90	70	91	91	27	40	65	86	52	49	16	19	67	67	17	54
	I2	60	58	19	25	36	47	42	25	47	40	37	51	54	59	59	64	90	100	37	42	41	39	95	72	92	93	36	46	63	89	45	48	10	13	61	61	19	57
	J1	38	39	40	45	71	43	39	57	41	73	42	36	29	26	45	45	43	37	100	34	76	76	33	33	35	43	56	79	37	34	46	42	29	29	50	50	24	38
	J2	42	49	14	19	56	44	65	25	58	38	34	43	48	49	45	45	40	42	34	100	50	35	39	42	39	43	33	50	54	44	57	49	4	5	52	52	19	46
	K1	42	45	22	33	77	41	43	72	46	78	56	41	38	38	47	46	46	41	76	50	100	83	40	42	40	49	74	85	47	44	44	39	17	19	63	63	18	30
	K2	42	39	28	34	57	34	35	75	38	78	79	38	30	34	39	44	37	39	76	35	83	100	33	34	31	41	85	77	38	37	31	34	21	25	56	56	20	20
	L1	53	52	13	23	33	46	40	21	47	34	28	50	52	58	58	57	90	95	33	39	40	33	100	74	96	90	28	40	63	90	44	47	6	8	58	58	17	59
	L2	75	53	12	23	37	47	41	23	45	37	31	51	56	53	66	67	70	72	33	42	42	34	74	100	69	68	31	43	79	79	47	49	5	7	59	60	18	56
	M1	50	52	13	24	34	48	32	20	42	33	28	42	51	52	56	53	91	92	35	39	40	31	96	69	100	88	27	40	61	89	46	48	6	9	61	61	18	51
	M2	58	56	22	29	42	49	46	35	50	45	36	50	56	59	61	66	91	93	43	43	49	41	90	68	88	100	36	42	70	90	46	45	16	20	66	66	17	59
	N1	39	36	15	33	46	27	30	79	28	82	84	27	30	45	28	38	27	36	56	33	74	85	28	31	27	36	100	67	31	34	21	30	9	15	41	41	16	13
	N2	50	48	22	28	80	36	44	54	44	72	50	37	38	39	42	46	40	46	79	50	85	77	40	43	40	42	67	100	41	43	38	41	11	14	55	55	22	31
	O1	77	56	19	28	43	52	53	31	58	37	32	47	68	68	71	69	65	63	37	54	47	38	63	79	61	70	31	41	100	80	44	40	17	19	58	58	16	64
	O2	68	61	15	23	36	49	41	24	47	35	35	45	58	61	57	60	86	89	34	44	44	37	90	79	89	90	34	43	80	100	43	45	8	10	61	61	18	61
	P1	50	72	24	28	42	73	64	22	61	35	22	75	46	34	53	47	52	45	46	57	44	31	44	47	46	46	21	38	44	43	100	87	17	14	62	61	21	47
	P2	45	73	19	29	35	68	55	19	54	35	34	73	44	33	49	48	49	48	42	49	39	34	47	49	48	45	30	41	40	45	87	100	7	11	56	55	22	39
	Q1	11	11	73	52	24	20	17	40	20	28	12	22	4	5	18	17	16	10	29	4	17	21	6	5	6	16	9	11	17	8	17	7	100	93	27	27	42	16
	Q2	13	13	65	60	23	19	17	44	21	32	17	22	7	9	21	21	19	13	29	5	19	25	8	7	9	20	15	14	19	10	14	11	93	100	26	26	38	23
S1	52	60	34	43	59	58	44	40	50	59	49	54	38	33	60	59	67	61	50	52	63	56	58	59	61	66	41	55	58	61	62	56	27	26	100	100	24	45	
S2	52	60	34	43	59	58	44	40	50	59	49	54	38	33	59	59	67	61	50	52	63	56	58	60	61	66	41	55	58	61	61	55	27	26	100	100	24	45	
R1	16	22	75	32	21	18	13	9	14	17	20	17	10	8	18	19	17	19	24	19	18	20	17	18	18	17	16	22	16	18	21	22	42	38	24	24	100	22	
R2	62	51	23	26	42	57	61	20	60	27	16	55	71	58	54	54	54	57	38	46	30	20	59	56	51	59	13	31	64	61	47	39	16	23	45	45	22	100	

Case 17 1971

Mrs C.W. is a 31 year old primigravida. She has essential hypertension, treated with a low dose of labetalol. She does not smoke. Her antenatal progress was good; she needed only a small increase in her dose of labetalol to maintain her BP at 120-140/90. Fetal growth was judged to be adequate both clinically and by ultrasound. However, at 39 weeks her BP was 150/100, and this was sustained during 48 hours as an inpatient. She had no proteinuria. Induction of labour was planned, but she laboured spontaneously at 39 weeks on 1/9/91. At 19.45hrs; cephalic 2/5. VE Cx 3cm dilated, effaced but thick. Station -2. ARM; clear liquor. FSE applied.

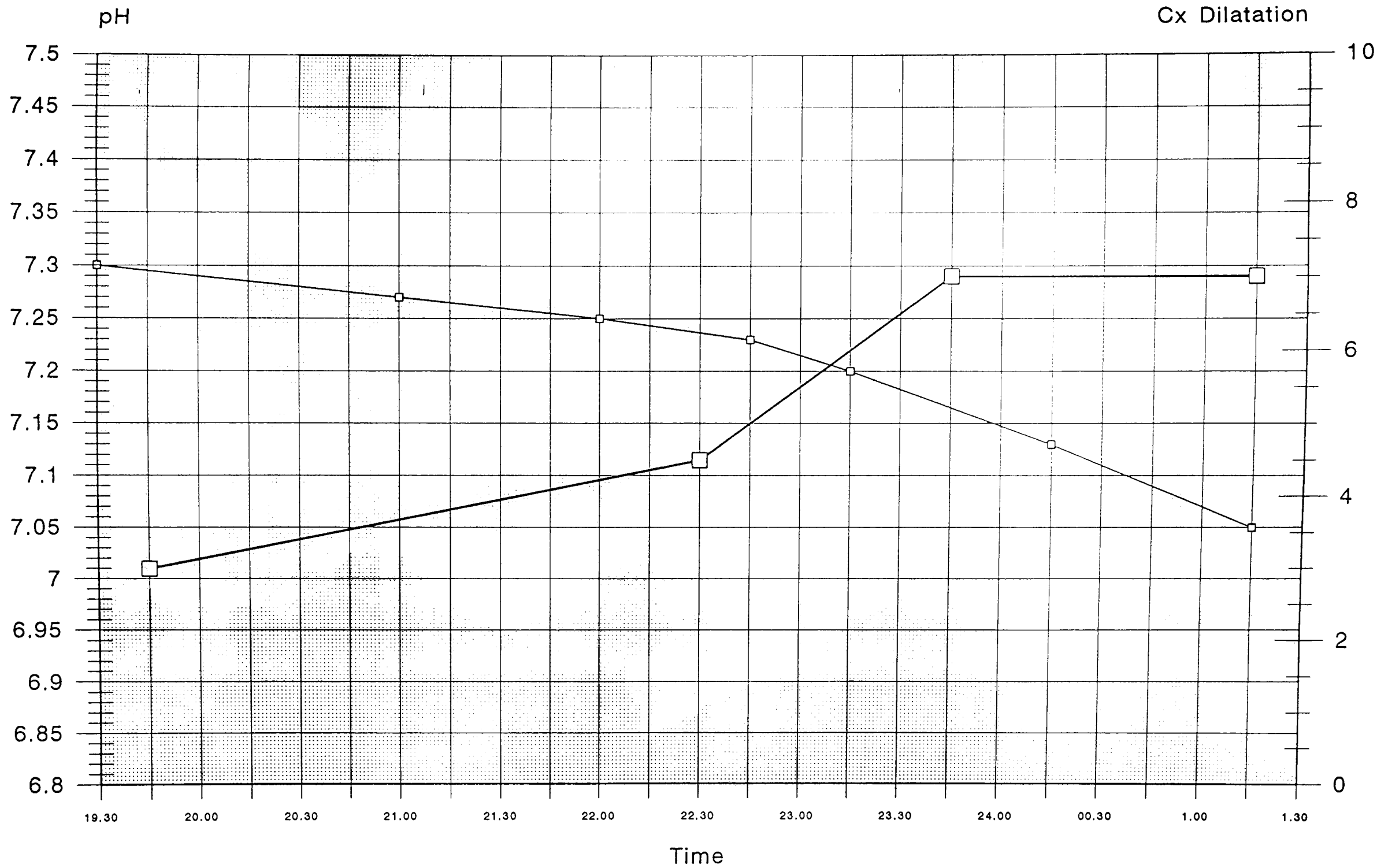
Labour events

Note; BP stable at 140/90 throughout labour.

- 20.20 Very frequent contractions; uterus scarcely relaxing.
- 21.00 FBS pH 7.27 BE +3
- 21.12 Epidural begun.
- 21.30 Epidural completed.
- 22.10 Maternal oxygen given. Position changed.
- 22.25 VE Cx 4-5cm dilated, thick. Station -2. Moulding ++.
- 22.50 FBS; pH 7.23 BE +3
- 23.00 Syntocinon begun.
- 23.50 VE Cx 7cm dilated, thin. Station -2.
- 00.15 FBS; pH 7.13 BE -2.
Decision for C/S.
- 00.35 Syntocinon stopped. Top up.
- 01.15 Male infant delivered by emergency C/S under epidural. Indication; low fetal scalp pH at 7cm dilatation.

Outcome

Birth weight 3.26kg
Apgar 6 & 9
Cord gases pH 7.21 BD(ecf) - / 1



○ pH □ Cx Dilatation (cm)

Agreement results matrix for case 17

		REVIEWER																																					
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	I1	I2	J1	J2	K1	K2	L1	L2	M1	M2	N1	N2	O1	O2	P1	P2	Q1	Q2	S1	S2	R1	R2
REVIEWER	A1	100	74	55	58	88	78	69	46	61	50	69	83	63	91	35	61	77	73	87	48	87	51	77	87	35	65	88	86	55	75	59	71	52	63	50	53	56	35
	A2	74	100	17	22	77	92	56	32	37	30	97	83	44	81	15	55	85	93	79	32	80	34	85	79	45	94	75	80	33	97	85	98	11	16	82	77	55	41
	B1	55	17	100	90	39	33	48	70	68	74	12	28	43	43	73	48	34	29	40	68	35	73	34	40	11	12	41	29	64	29	5	16	95	86	4	5	34	27
	B2	58	22	90	100	48	37	70	78	82	82	18	35	48	46	56	66	38	30	49	80	46	84	38	49	17	18	48	39	79	30	9	20	90	91	9	9	36	31
	C1	88	77	39	48	100	84	81	48	65	53	74	79	57	89	28	81	91	79	95	49	96	53	91	95	35	63	96	91	62	81	59	74	31	49	51	68	58	42
	C2	78	92	33	37	84	100	60	34	45	41	90	79	49	84	33	65	95	92	84	40	80	45	95	84	41	86	75	74	36	93	79	92	25	28	77	72	64	43
	D1	69	56	48	70	81	60	100	74	94	79	52	53	49	70	30	92	75	52	85	81	81	81	75	85	33	34	79	81	87	55	23	52	40	65	15	52	43	38
	D2	46	32	70	78	48	34	74	100	85	96	28	40	37	42	51	72	38	39	46	91	52	92	38	46	26	29	53	46	93	40	14	30	57	60	15	16	36	40
	E1	61	37	68	82	65	45	94	85	100	89	32	48	45	59	40	87	59	35	68	92	64	92	59	68	30	32	64	64	90	38	13	33	58	72	12	34	40	34
	E2	50	30	74	82	53	41	79	96	89	100	26	38	39	47	58	77	45	39	52	96	52	96	45	52	24	27	52	43	89	38	13	27	60	63	14	15	38	40
	F1	69	97	12	18	74	90	52	28	32	26	100	83	42	80	11	51	83	92	78	28	77	28	83	78	43	93	73	76	28	92	87	98	7	12	87	83	52	39
	F2	83	83	28	35	79	79	53	40	48	38	83	100	59	90	16	41	68	79	83	40	82	41	68	83	41	85	83	83	48	81	80	83	22	34	73	58	54	36
	G1	63	44	43	48	57	49	49	37	45	39	42	59	100	63	27	40	49	48	57	35	56	37	49	57	69	43	58	56	45	49	40	41	46	54	32	28	57	50
	G2	91	81	43	46	89	84	70	42	59	47	80	90	63	100	29	59	80	79	93	48	89	48	80	93	36	76	87	89	50	78	70	79	37	48	62	67	57	35
	H1	35	15	73	56	28	33	30	51	40	58	11	16	27	29	100	35	34	26	27	52	21	58	34	27	11	12	24	17	33	24	9	16	62	42	10	11	34	26
	H2	61	55	48	66	81	65	92	72	87	77	51	41	40	59	35	100	82	54	75	78	76	81	82	75	30	31	71	72	83	58	27	51	37	56	19	61	51	42
	I1	77	85	34	38	91	95	75	38	59	45	83	68	49	80	34	82	100	90	86	39	86	45	100	86	38	71	82	79	53	89	67	85	25	39	63	76	61	46
	I2	73	93	29	30	79	92	52	39	35	39	92	79	48	79	26	54	90	100	78	31	82	35	90	78	41	88	79	74	39	97	83	94	22	25	82	77	64	45
	J1	87	79	40	49	95	84	85	46	68	52	78	83	57	93	27	75	86	78	100	53	96	53	86	100	38	67	94	93	63	77	59	79	33	50	53	67	53	36
	J2	48	32	68	80	49	40	81	91	92	96	28	40	35	48	52	78	39	31	53	100	49	98	39	53	26	29	47	46	85	32	14	30	55	58	15	16	33	31
	K1	87	80	35	46	96	80	81	52	64	52	77	82	56	89	21	76	86	82	96	49	100	50	86	96	39	67	99	94	66	82	59	78	29	48	51	68	59	44
	K2	51	34	73	84	53	45	81	92	92	96	28	41	37	48	58	81	45	35	53	98	50	100	45	53	26	29	50	46	87	38	13	31	60	62	14	15	39	36
	L1	77	85	34	38	91	95	75	38	59	45	83	68	49	80	34	82	100	90	86	39	86	45	100	86	38	71	82	79	53	89	67	85	25	39	63	76	61	46
	L2	87	79	40	49	95	84	85	46	68	52	78	83	57	93	27	75	86	78	100	53	96	53	86	100	38	67	94	93	63	77	59	79	33	50	53	67	53	36
	M1	35	45	11	17	35	41	33	26	30	24	43	41	69	36	11	30	38	41	38	26	39	26	38	38	100	45	36	36	26	43	36	43	6	11	35	25	44	59
	M2	65	94	12	18	63	86	34	29	32	27	93	85	43	76	12	31	71	88	67	29	67	29	71	67	45	100	63	65	28	89	91	94	7	11	92	68	52	43
	N1	88	75	41	48	96	75	79	53	64	52	73	83	58	87	24	71	82	79	94	47	99	50	82	94	36	63	100	95	70	79	57	73	34	53	48	65	55	40
	N2	86	80	29	39	91	74	81	46	64	43	76	83	56	89	17	72	79	74	93	46	94	46	79	93	36	65	95	100	64	78	60	77	23	44	49	69	51	34
	O1	55	33	64	79	62	36	87	93	90	89	28	48	45	50	33	83	53	39	63	85	66	87	53	63	26	28	70	64	100	40	15	30	53	68	14	37	43	38
	O2	75	97	29	30	81	93	55	40	38	38	92	81	49	78	24	58	89	97	77	32	82	38	89	77	43	89	79	78	40	100	81	95	21	25	77	72	62	49
	P1	59	85	5	9	59	79	23	14	13	13	87	80	40	70	9	27	67	83	59	14	59	13	67	59	36	91	57	60	15	81	100	88	0	1	95	74	57	32
	P2	71	98	16	20	74	92	52	30	33	27	98	83	41	79	16	51	85	94	79	30	78	31	85	79	43	94	73	77	30	95	88	100	10	14	87	82	57	35
Q1	52	11	95	90	31	25	40	57	58	60	7	22	46	37	62	37	25	22	33	55	29	60	25	33	6	7	34	23	53	21	0	10	100	92	0	0	26	21	
Q2	63	16	86	91	49	28	65	60	72	63	12	34	54	48	42	56	39	25	50	58	48	62	39	50	11	11	53	44	68	25	1	14	92	100	0	12	29	24	
S1	50	82	4	9	51	77	15	15	12	14	87	73	32	62	10	19	63	82	53	15	51	14	63	53	35	92	48	49	14	77	95	87	0	0	100	75	51	36	
S2	53	77	5	9	68	72	52	16	34	15	83	58	28	67	11	61	76	77	67	16	68	15	76	67	25	68	65	69	37	72	74	82	0	12	75	100	48	30	
R1	56	55	34	36	58	64	43	36	40	38	52	54	57	57	34	51	61	64	53	33	59	39	61	53	44	52	55	51	43	62	57	57	26	29	51	48	100	67	
R2	35	41	27	31	42	43	38	40	34	40	39	36	50	35	26	42	46	45	36	31	44	36	46	36	59	43	40	34	38	49	32	35	21	24	36	30	67	100	

Case 18 959

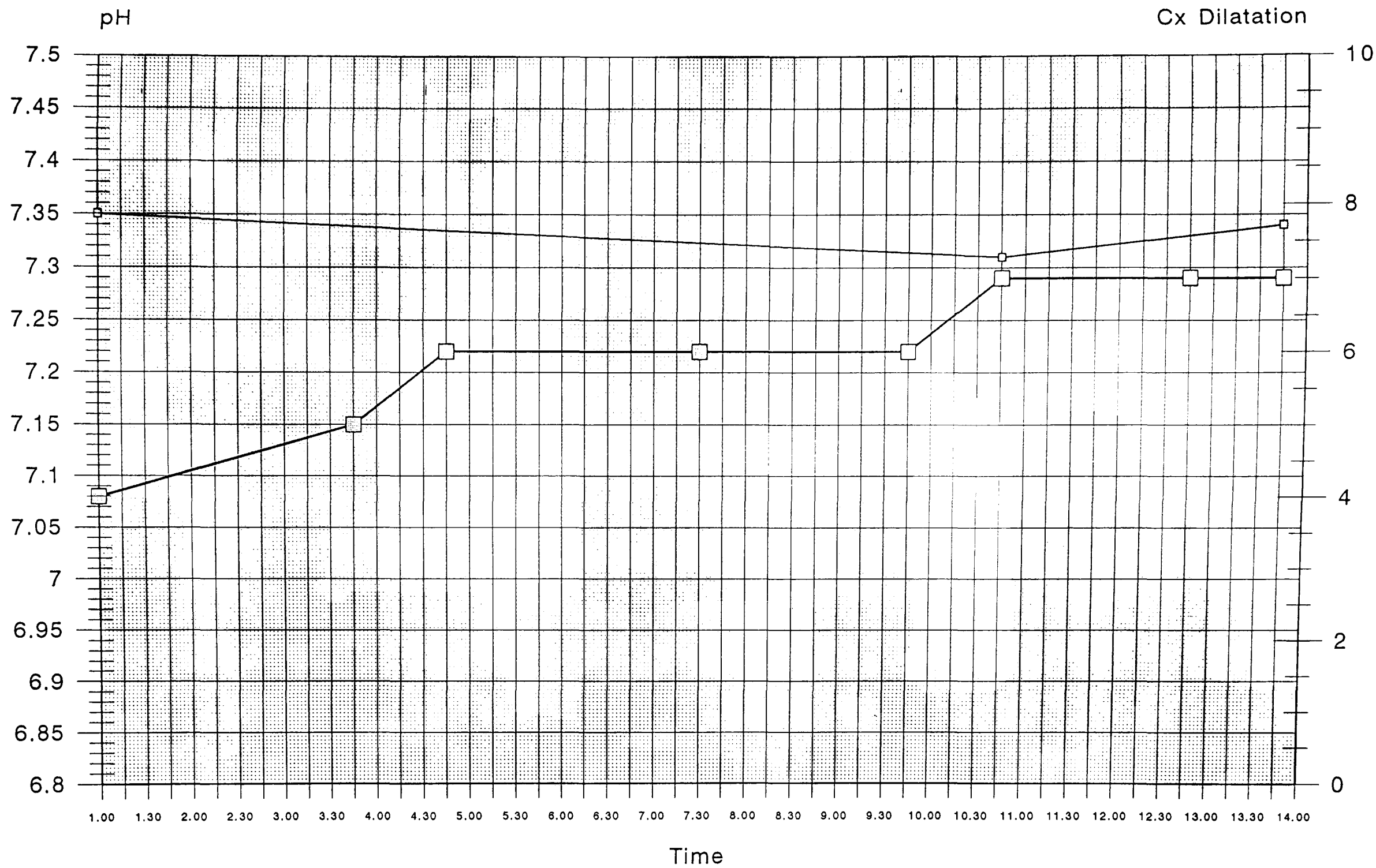
Mrs K.D. is a 31 year old lady expecting her first baby. She has had one early miscarriage. She is a fit lady and a non smoker. Her pregnancy was straightforward, and she laboured spontaneously at 41 weeks gestation. She was admitted on 19.1.91 in early labour; presentation cephalic 3/5, and the cervix at 22.35hrs was 3cm dilated, part effaced. Intermittent monitoring was performed. By 01.00hrs the cervix was 4cm dilated, effaced and thin. ARM was performed, with the drainage of a small amount of clear liquor.

Labour events

- 03.50 VE Cx 5cm dilated, thin. Station -1. OP position. FSE applied.
- 04.30 Epidural inserted.
- 04.50 VE Cx 6cm dilated. Station -1. OP position.
- 06.30 Syntocinon started.
- 07.30 VE Cx 6cm dilated. Station -1. OP position.
- 07.40 Top up.
- 09.40 VE No change.
- 10.40 VE Cx 7cm dilated. Station -1. Direct OP position. IUPD inserted.
FBS pH 7.31 BE -0.9
- 12.40 VE No change. Decision for LSCS. Syntocinon stopped.
- 13.41 Male infant delivered by C/S under epidural. Indication; "failure to progress".
Thick meconium noted at operation.

Outcome

Birth weight 3.4kg
Apgar 9 & 9
Cord gases pH 7.34 / 7.35 BD(ecf) - / 3



pH
 Cx Dilatation (cm)

Agreement results matrix for case 18

		REVIEWER																																					
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	I1	I2	J1	J2	K1	K2	L1	L2	M1	M2	N1	N2	O1	O2	P1	P2	Q1	Q2	S1	S2	R1	R2
REVIEWER	A1	100	80	51	55	85	89	86	83	89	81	67	69	79	79	86	81	76	82	84	49	89	77	84	71	60	62	58	60	82	82	72	63	41	43	73	73	57	47
	A2	80	100	46	47	87	84	84	81	80	82	85	70	86	86	79	81	82	75	84	47	86	88	82	85	69	72	80	69	78	77	88	80	35	28	67	67	53	56
	B1	51	46	100	89	49	45	45	37	44	38	42	36	38	41	46	36	41	39	46	42	45	41	44	37	16	17	32	16	46	41	35	31	68	70	23	22	27	20
	B2	55	47	89	100	48	49	44	43	44	43	40	42	37	38	54	41	41	48	52	46	49	40	49	42	20	21	31	21	46	45	37	32	64	75	29	29	42	36
	C1	85	87	49	48	100	86	92	84	81	80	77	73	82	82	80	84	79	78	77	44	85	85	84	79	70	72	74	69	80	79	81	74	34	36	71	71	57	52
	C2	89	84	45	49	86	100	88	86	80	84	80	75	79	79	85	84	75	82	75	50	97	84	88	78	70	71	70	70	81	84	79	72	37	38	69	68	60	54
	D1	86	84	45	44	92	88	100	87	88	87	73	76	88	86	80	86	82	80	80	46	91	89	81	74	76	79	70	75	77	81	88	79	33	33	74	73	59	55
	D2	83	81	37	43	84	86	87	100	89	96	69	76	79	77	84	82	77	90	80	47	87	79	87	73	76	79	65	76	78	81	84	77	30	33	84	84	64	54
	E1	89	80	44	44	81	80	88	89	100	89	67	74	87	87	85	73	86	89	85	43	81	76	81	63	72	75	66	73	84	81	83	74	36	33	84	84	55	47
	E2	81	82	38	43	80	84	87	96	89	100	65	72	76	76	81	81	74	89	80	46	85	76	90	76	73	76	67	73	80	80	81	71	28	33	81	81	65	57
	F1	67	85	42	40	77	80	73	69	67	65	100	61	78	76	70	73	72	63	70	44	78	86	79	78	74	74	87	73	81	71	83	88	35	20	63	62	53	55
	F2	69	70	36	42	73	75	76	76	74	72	61	100	72	69	66	61	67	71	67	35	73	76	65	47	70	70	52	72	66	66	76	63	24	41	69	69	58	43
	G1	79	86	38	37	82	79	88	79	87	76	78	72	100	98	78	80	90	76	82	45	82	89	67	76	77	80	75	76	64	88	93	85	34	22	73	72	45	52
	G2	79	86	41	38	82	79	86	77	87	76	76	69	98	100	76	75	91	78	80	45	79	83	72	77	70	73	77	70	68	89	89	82	35	22	67	67	44	52
	H1	86	79	46	54	80	85	80	84	85	81	70	66	78	76	100	81	80	89	81	53	87	73	83	72	65	67	68	66	76	79	74	66	40	37	73	73	59	58
	H2	81	81	36	41	84	84	86	82	73	81	73	61	80	75	81	100	67	71	76	46	87	82	89	91	81	83	79	81	78	79	69	70	34	30	77	77	62	58
	I1	76	82	41	41	79	75	82	77	86	74	72	67	90	91	80	67	100	76	78	52	76	78	70	67	65	68	72	66	75	83	87	80	36	23	68	68	46	56
	I2	82	75	39	48	78	82	80	90	89	89	63	71	76	78	89	71	76	100	85	54	81	70	85	66	65	68	61	66	76	80	76	69	33	34	75	75	69	63
	J1	84	84	46	52	77	75	80	80	85	80	70	67	82	80	81	76	78	85	100	50	76	82	74	71	64	67	62	64	77	75	78	71	39	32	75	75	68	64
	J2	49	47	42	46	44	50	46	47	43	46	44	35	45	45	53	46	52	54	50	100	49	44	51	48	30	33	33	32	42	51	44	46	57	30	30	30	43	49
	K1	89	86	45	49	85	97	91	87	81	85	78	73	82	79	87	87	76	81	76	49	100	87	87	78	73	76	73	72	78	83	83	75	36	37	70	69	59	53
	K2	77	88	41	40	85	84	89	79	76	76	86	76	89	83	73	82	78	70	82	44	87	100	73	75	79	80	76	78	74	77	91	86	34	27	72	70	57	54
	L1	84	82	44	49	84	88	81	87	81	90	79	65	67	72	83	89	70	85	74	51	87	73	100	87	72	75	76	72	85	75	70	74	36	35	75	74	63	57
	L2	71	85	37	42	79	78	74	73	63	76	78	47	76	77	72	91	67	66	71	48	78	75	87	100	70	73	87	71	75	80	68	73	33	22	63	63	58	62
	M1	60	69	16	20	70	70	76	76	72	73	74	70	77	70	65	81	65	65	64	30	73	79	72	70	100	98	82	96	71	65	81	82	16	13	89	86	54	52
	M2	62	72	17	21	72	71	79	79	75	76	74	70	80	73	67	83	68	68	67	33	76	80	75	73	98	100	84	95	73	65	82	84	18	13	88	86	55	54
	N1	58	80	32	31	74	70	70	65	66	67	87	52	75	77	68	79	72	61	62	33	73	76	76	87	82	84	100	79	74	69	77	84	24	13	69	67	48	53
	N2	60	69	16	21	69	70	75	76	73	73	73	72	76	70	66	81	66	66	64	32	72	78	72	71	96	95	79	100	74	66	79	80	18	14	88	88	55	54
	O1	82	78	46	46	80	81	77	78	84	80	81	66	64	68	76	78	75	76	77	42	78	74	85	75	71	73	74	74	100	60	69	74	36	32	81	82	63	53
	O2	82	77	41	45	79	84	81	81	81	80	71	66	88	89	79	79	83	80	75	51	83	77	75	80	65	65	69	66	60	100	82	74	37	30	67	67	52	57
	P1	72	88	35	37	81	79	88	84	83	81	83	76	93	89	74	69	87	76	78	44	83	91	70	68	81	82	77	79	69	82	100	92	28	21	74	72	53	54
	P2	63	80	31	32	74	72	79	77	74	71	88	63	85	82	66	70	80	69	71	46	75	86	74	73	82	84	84	80	74	74	92	100	29	13	69	67	57	57
Q1	41	35	68	64	34	37	33	30	36	28	35	24	34	35	40	34	36	33	39	57	36	34	36	33	16	18	24	18	36	37	28	29	100	63	22	23	30	20	
Q2	43	28	70	75	36	38	33	33	33	33	20	41	22	22	37	30	23	34	32	30	37	27	35	22	13	13	13	14	32	30	21	13	63	100	22	22	44	13	
S1	73	67	23	29	71	69	74	84	84	81	63	69	73	67	73	77	68	75	75	30	70	72	75	63	89	88	69	88	81	67	74	69	22	22	100	99	58	48	
S2	73	67	22	29	71	68	73	84	84	81	62	69	72	67	73	77	68	75	75	30	69	70	74	63	86	86	67	88	82	67	72	67	23	22	99	100	58	48	
R1	57	53	27	42	57	60	59	64	55	65	53	58	45	44	59	62	46	69	68	43	59	57	63	58	54	55	48	55	63	52	53	57	30	44	58	58	100	78	
R2	47	56	20	36	52	54	55	54	47	57	55	43	52	52	58	58	56	63	64	49	53	54	57	62	52	54	53	54	53	57	54	57	20	13	48	48	78	100	

Case 19 1887

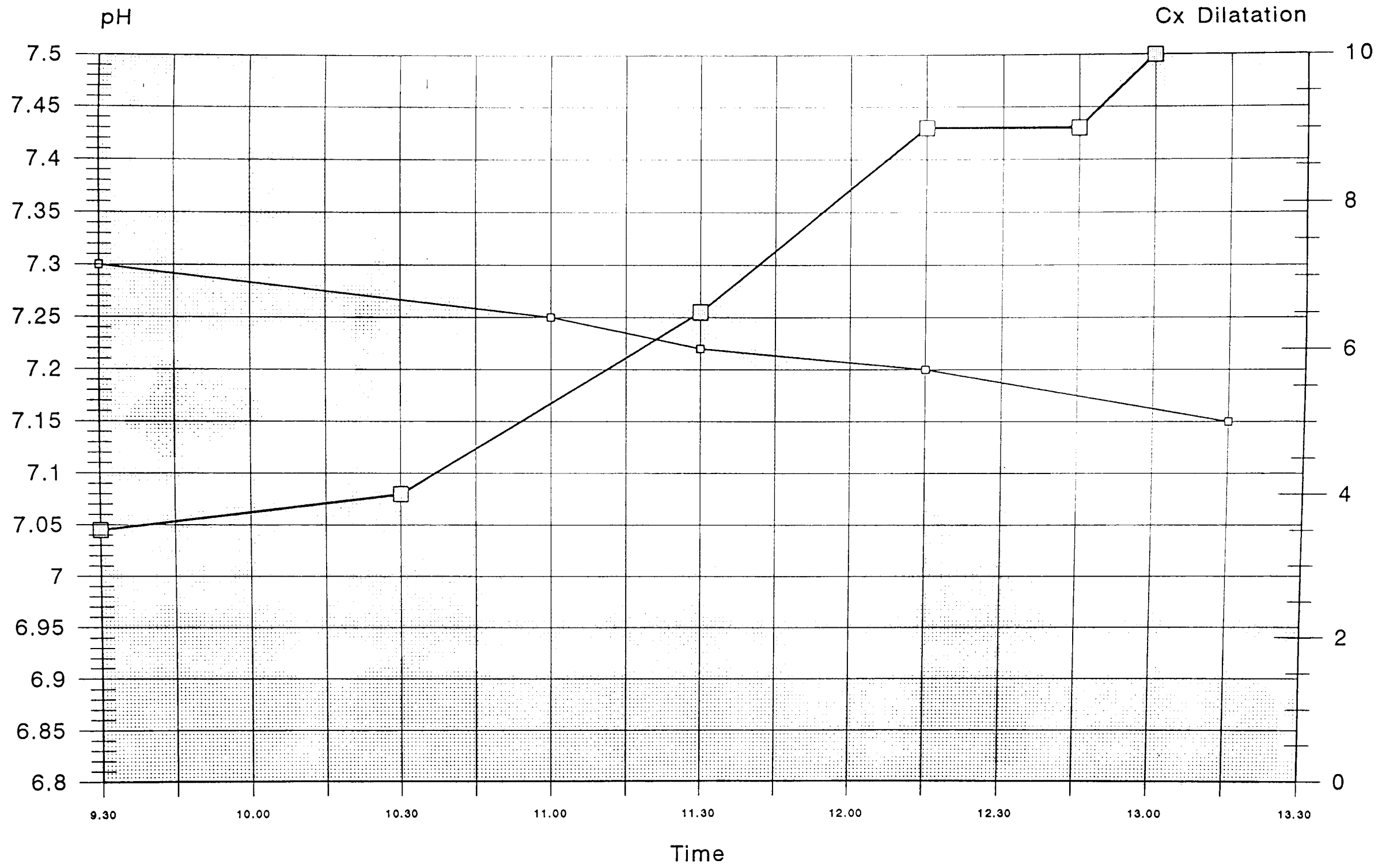
Miss L.B. is an 18 year old girl expecting her second baby. She is fit and well, but smokes 10 cigarettes per day. Her first child was delivered by emergency Caesarean section because of fetal distress which occurred at 4cm dilatation in a term spontaneous labour. The baby (girl) weighed 2.44kg. During the current pregnancy fetal growth was judged clinically to be satisfactory, and ultrasound measurements at 30 weeks were well within the normal range. There was no hypertension or proteinuria. Xray pelvimetry demonstrated good pelvic capacity and a trial of vaginal delivery was planned. Spontaneous labour occurred at 37 weeks; when she was admitted on 20.8.91 the fetal head was 3/5 palpable and the cervix 2cm dilated (03.00hrs).

Labour events

- 09.50 Contractions more painful. Continuous monitoring begun.
- 10.30 VE Cx 4cm dilated, fully effaced. Station -1. ARM; clear liquor. FSE applied.
- 10.50 Epidural inserted.
- 11.00 Oxygen given.
- 11.35 VE Cx 6-7cm dilated. Station +1. LOT position.
FBS; pH 7.22 BE -3
- 12.20 VE Cx 9cm dilated. Station +1. LOT position.
FBS; pH 7.20 BE -4
- 12.45 VE Cx 9cm dilated.
- 13.00 VE Cx fully dilated. Station +1. LOT position. Caput + moulding +.
- 13.22 Male infant delivered by Ventouse extraction. Indication; continuing fetal heart decelerations and borderline scalp pH.

Outcome

Birth weight 2.30kg
Apgar 9 & 9
Cord gases pH 7.15 / 7.23 BD(ecf) 5 / 3



○ pH □ Cx Dilatation (cm)

Case 20 1558

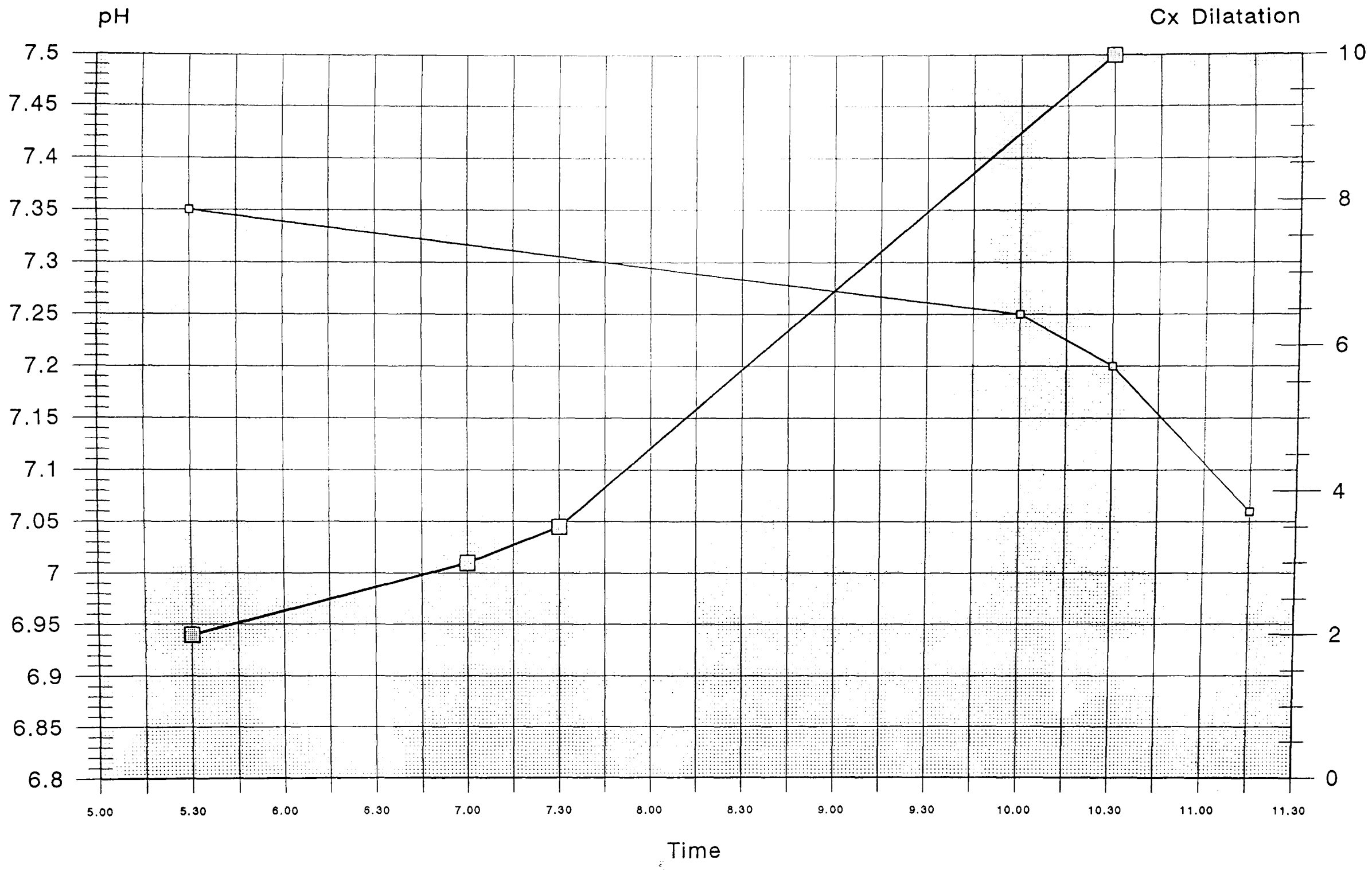
Mrs J.D. is a 32 year old primigravida; a fit lady, and a nonsmoker. Her pregnancy was straightforward until she was admitted at 35 weeks with spontaneous rupture of the membranes. After 36 hours labour had still not established. The uterus was thought to be small for dates (fundal height 31cm); cephalic 2/5. VE Cx closed, thick. Clear liquor was seen. Labour was induced with vaginal prostaglandins, and on 29/5/91 at 09.50hrs prostin gel 2mg PV was given. This was repeated 11 hours later; VE being unchanged. After a further 10 hours (05.00hrs on 30/5/91) she had regular contractions, and the cervix was 2cm dilated, effacing; station -1.

Labour events

05.35 VE Cx 2cm dilated. FSE applied.
07.00 VE Cx 3cm dilated, effaced. Station 0.
07.30 VE Cx 3-4cm dilated. Station -1.
07.40 Epidural begun.
08.15 Epidural completed.
09.00 Syntocinon started.
09.50 Top up.
10.25 VE Cx fully dilated. Station +1.
10.40 Pushing begun.
11.15 Normal delivery of female infant.

Outcome

Birth weight 1.66kg
Apgar 6 & 9
Cord gases pH 7.06 / 7.09 BD(ecf) 3 / 5



□ pH ■ Cx Dilatation (cm)

Case 21 1932

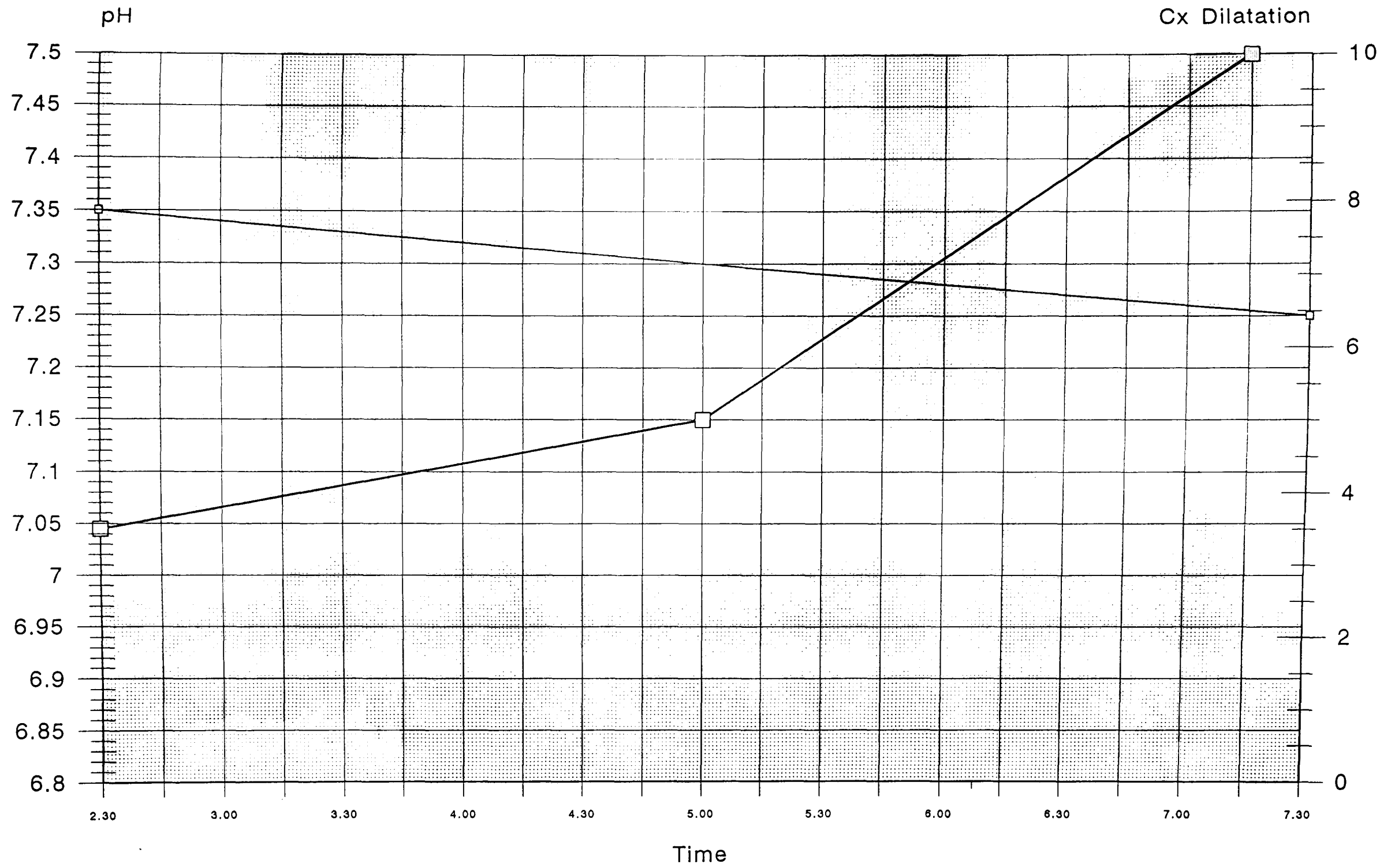
Mrs T.D. is a 26 year old lady expecting her fourth child. She is a fit lady, but smokes 15 cigarettes per day. Her first two children were born normally at term and weighed 6lb 7oz and 7lb 3oz respectively. Her third child was delivered by Caesarean section because of fetal distress during preterm labour at 34 weeks. The baby was appropriately grown. Her current pregnancy progressed normally and a trial of vaginal delivery was planned. She was admitted in spontaneous labour at 40 weeks, on 25.8.91. The uterus was term size, presentation cephalic 2/5. VE (at 02.35hrs) Cx 3-4cm dilated, effaced and thin. Station -1. Clear liquor seen.

Labour events

05.00 VE Cx 5cm dilated. Station -1. FSE applied.
05.30 Epidural begun. (Not effective)
06.30 Top up.
07.10 VE Cx fully dilated. Vertex visible.
07.20 Pushing begun.
07.28 Normal delivery of male infant.

Outcome

Birth weight 3.60kg
Apgar 9 & 9
Cord gases pH 7.25 / 7.42 BD(ecf) 10 / 2



○ pH □ Cx Dilatation (cm)

Case 22 237

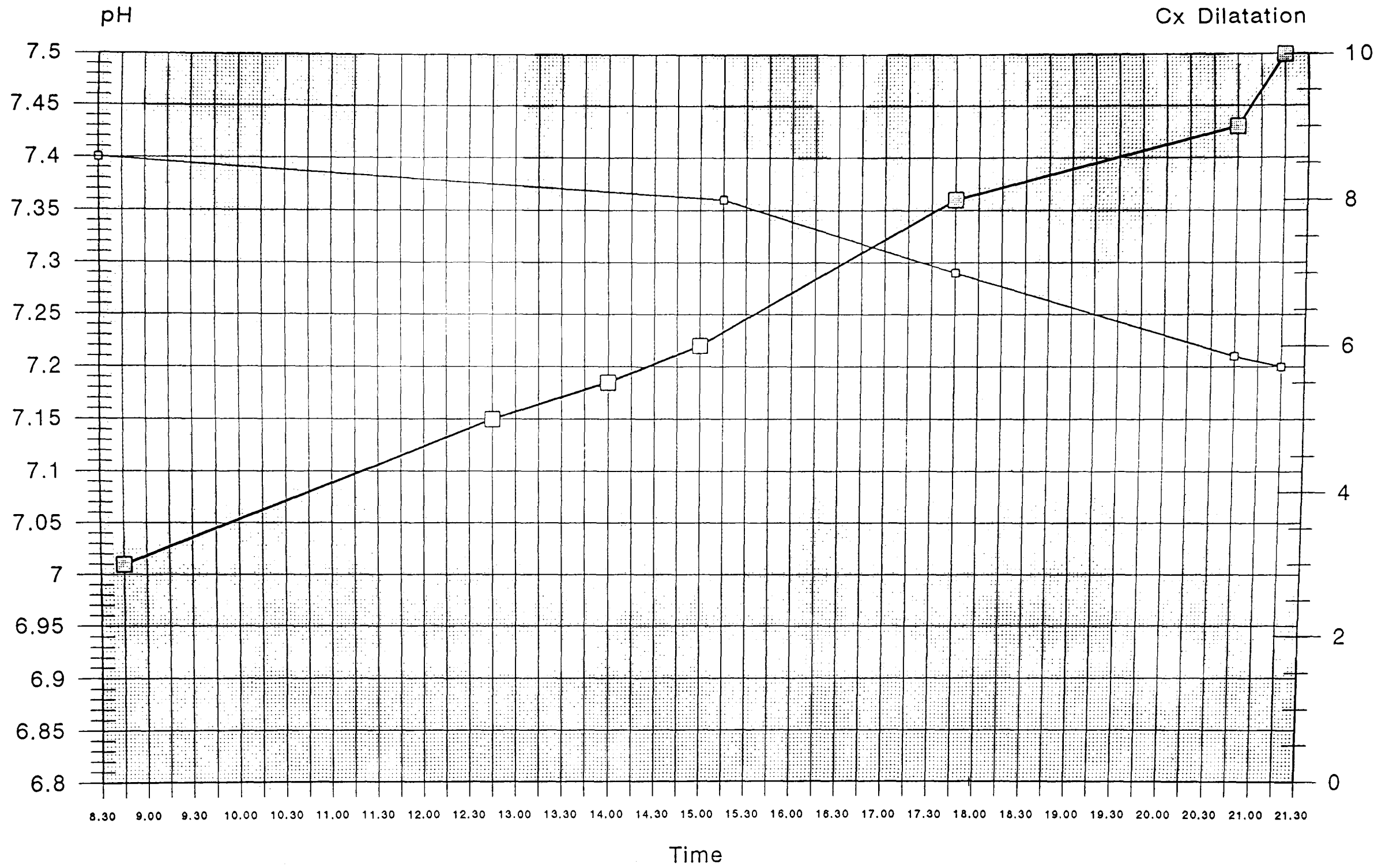
Mrs C.H. is a 26 year old primigravida; a fit lady, and a nonsmoker. Her pregnancy was straightforward and she laboured spontaneously at 41 weeks. She was admitted on 17/8/90 at 08.50hrs in early labour. Cephalic 3/5. VE Cx 3cm dilated, 50% effaced. Station -2.

Labour events

- 10.00 SR0M; clear liquor.
- 10.15 Pethidine 100mg, phenergan 25mg im.
- 12.40 VE Cx 5cm dilated, fully effaced & thin. Station -1. FSE applied.
- 13.30 Epidural.
- 13.50 Change of maternal position.
- 13.55 VE Cx 5-6cm dilated. Station -1. LOA position.
- 14.30 Hypotensive (BP 90/60). IV fluids increased.
- 14.55 VE Cx 6cm dilated. Station -1.
- 15.20 FBS; pH 7.36 BE -6
- 15.45 Top up. Syntocinon started.
- 17.45 VE Cx 8cm dilated. Station 0.
FBS; pH 7.29 BE -4
- 18.05 Top up.
- 20.00 Top up.
- 20.45 VE Cx 9cm dilated. Station +1. OA position.
FBS; pH 7.21 BE -3.5
- 21.15 VE Cx fully dilated. Head low & descending well with pushing.
- 21.20 Normal delivery of male infant.

Outcome

Birth weight 3.54kg
Apgar 5 & 9
Cord gases pH 7.20 / 7.24 BD(ecf) 4 / 6



○ pH □ Cx Dilatation (cm)

Case 23 1222

Mrs M.D. is a 29 year old primigravida. She is a fit lady and does not smoke. She required several admissions for hyperemesis in the first 16 weeks, and continued to be troubled with vomiting throughout the whole of her pregnancy. However, she gained weight normally and fetal growth was judged to be adequate both on clinical assessment and by ultrasound. She was admitted in spontaneous labour at 42 weeks on 21/3/91 at 23.30hrs. Uterus term size. Cephalic 3/5. VE (at 00.15hrs) Cx 2cm dilated, fully effaced & thin. Station -2.

Labour events

Note; mild maternal pyrexia from 03.30 to 07.00 (37.5 - 37.8°C).

- 01.50 VE Cx 3cm dilated, thin.
- 02.05 Epidural begun.
- 02.40 Epidural complete.
- 03.00 VE Cx 5cm dilated. Station -2. ARM; no liquor seen. Caput ++.
- 05.00 VE Cx 5cm dilated. Station -2.
- 05.20 Hypotensive (BP 75/40). IV fluids increased.
- 05.30 Top up.
- 06.00 Vomiting.
- 06.30 Vomiting.
- 07.45 VE Cx 5cm dilated. Station -2. ROP position.
- 08.00 Syntocinon started. Top up. Maternal hypotension during top up (BP 90/55). IV fluids increased.
- 09.15 Vomiting.
- 10.10 Top up.
- 11.10 Vomiting.
- 11.20 VE Cx 9cm dilated. Station -1. Direct OP position. Caput +. FSE reapplied. Syntocinon stopped.
- 11.30 Top up.
- 11.55 FBS; pH 7.35 BE -2
- 12.00 Syntocinon started.
- 13.30 Vomiting.
- 13.35 VE Cx 8-9cm dilated, thick. Direct OP position. Station -1.
- 14.00 Decision for C/S.
- 14.15 Top up.
- 15.17 Male delivered by emergency C/S under epidural. Indication; failure to progress in first stage of labour; OP position. Thick meconium noted at operation.

Outcome

Birth weight 3.56kg

Apgar 8 & 9

Cord gases pH 7.30 / 7.37 BD(ecf) 0 / 0

Case 23 1222

Mrs M.D. is a 29 year old primigravida. She is a fit lady and does not smoke. She required several admissions for hyperemesis in the first 16 weeks, and continued to be troubled with vomiting throughout the whole of her pregnancy. However, she gained weight normally and fetal growth was judged to be adequate both on clinical assessment and by ultrasound. She was admitted in spontaneous labour at 42 weeks on 21/3/91 at 23.30hrs. Uterus term size. Cephalic 3/5. VE (at 00.15hrs) Cx 2cm dilated, fully effaced & thin. Station -2.

Labour events

Note; mild maternal pyrexia from 03.30 to 07.00 (37.5 - 37.8°C).

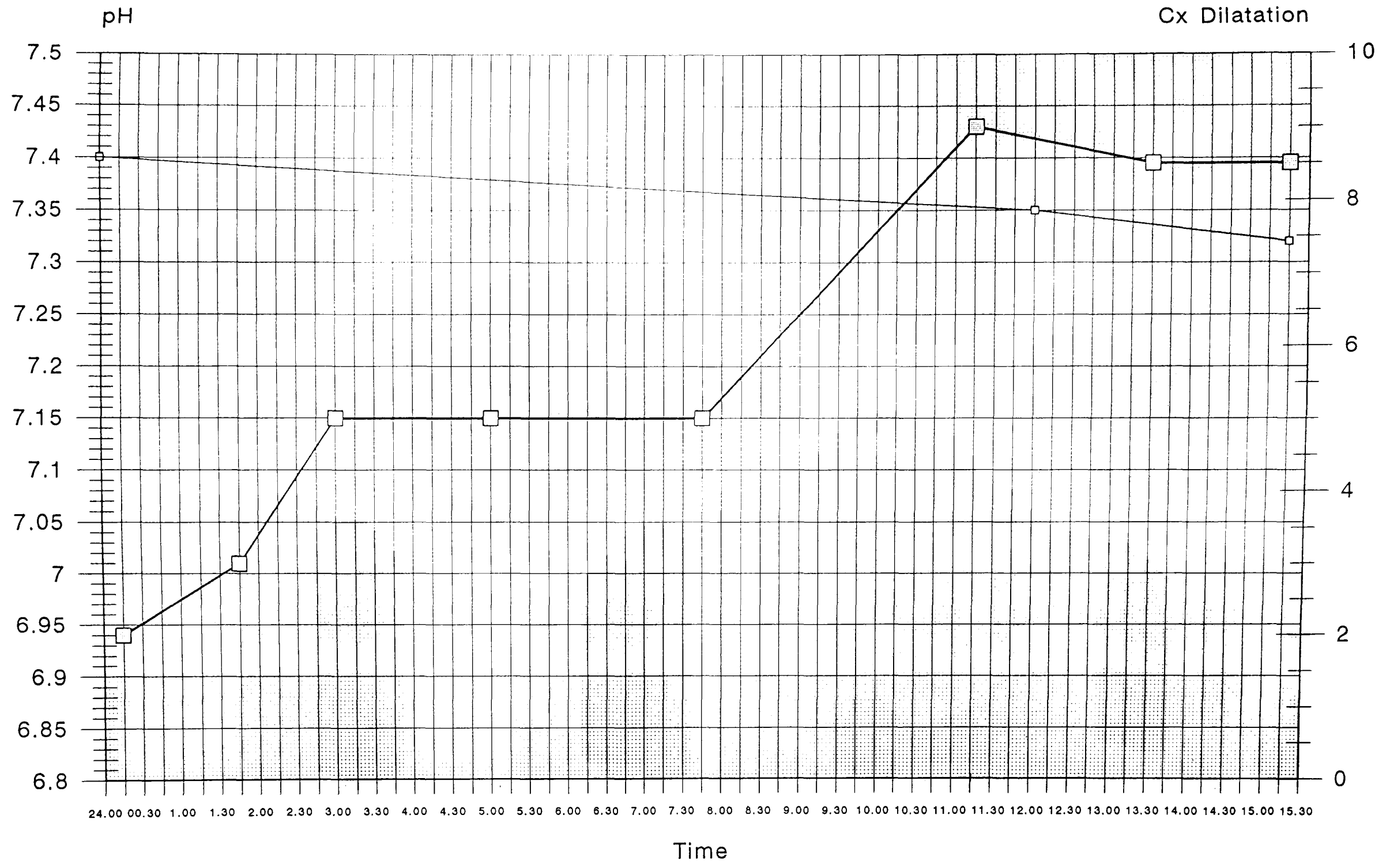
- 01.50 VE Cx 3cm dilated, thin.
- 02.05 Epidural begun.
- 02.40 Epidural complete.
- 03.00 VE Cx 5cm dilated. Station -2. ARM; no liquor seen. Caput ++.
- 05.00 VE Cx 5cm dilated. Station -2.
- 05.20 Hypotensive (BP 75/40). IV fluids increased.
- 05.30 Top up.
- 06.00 Vomiting.
- 06.30 Vomiting.
- 07.45 VE Cx 5cm dilated. Station -2. ROP position.
- 08.00 Syntocinon started. Top up. Maternal hypotension during top up (BP 90/55). IV fluids increased.
- 09.15 Vomiting.
- 10.10 Top up.
- 11.10 Vomiting.
- 11.20 VE Cx 9cm dilated. Station -1. Direct OP position. Caput +. FSE reapplied. Syntocinon stopped.
- 11.30 Top up.
- 11.55 FBS; pH 7.35 BE -2
- 12.00 Syntocinon started.
- 13.30 Vomiting.
- 13.35 VE Cx 8-9cm dilated, thick. Direct OP position. Station -1.
- 14.00 Decision for C/S.
- 14.15 Top up.
- 15.17 Male delivered by emergency C/S under epidural. Indication; failure to progress in first stage of labour; OP position. Thick meconium noted at operation.

Outcome

Birth weight 3.56kg

Apgar 8 & 9

Cord gases pH 7.30 / 7.37 BD(ecf) 0 / 0



□ pH ■ Cx Dilatation (cm)

Case 24 1094

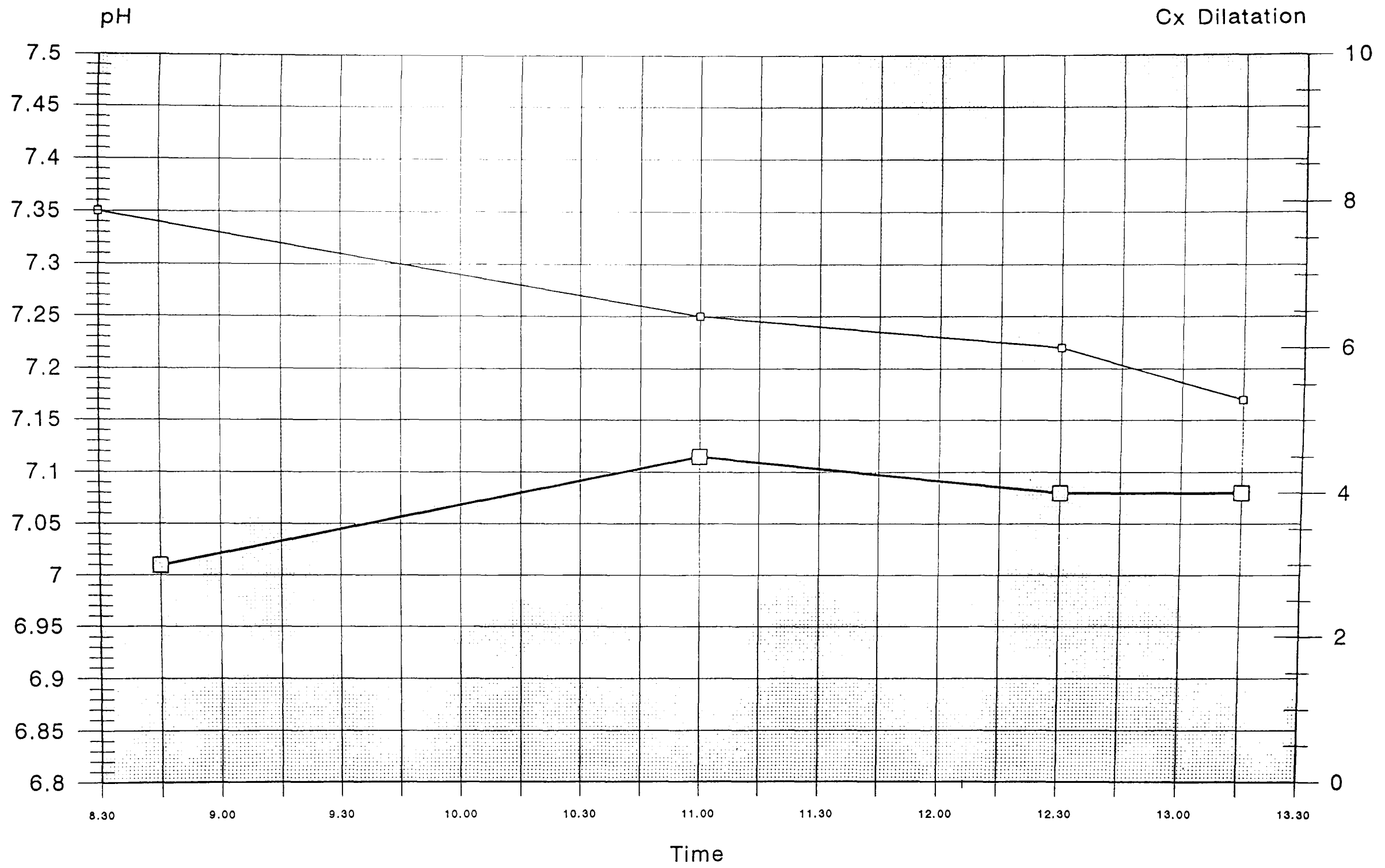
Miss J.D. is a 19 year old primigravida. She is fit and well, and does not smoke. There were no antenatal problems, and spontaneous labour occurred at 41 weeks. She was admitted on 16/2/91 at 08.20hrs in early labour. Her BP was raised at 160/95 and there was 2+ proteinuria. Miss J.D. was asymptomatic and reflexes were normal. The baby was judged to be well grown; cephalic 3/5. VE (at 08.50hrs) Cx 3cm dilated, effaced. Station -3.

Labour events

- 09.30 In the bath.
- 10.30 Pethidine 100mg, Phenergan 25mg im.
- 10.45 Vomiting.
- 10.50 Maternal oxygen given.
- 10.55 VE Cx 4-5cm dilated, thin. Station -2. ARM; clear liquor. FSE applied.
- 11.40 Unsuccessful attempt at FBS. (Head easily pushed out of pelvis)
- 12.15 Ranitidine 50mg, maxolon 10mg iv.
- 12.25 VE Cx 4cm dilated. FBS pH 7.22 BE -13
- 12.35 Decision for C/S.
- 13.22 Male infant delivered by emergency C/S under GA. Indication; prolonged decelerations at 4cm dilatation.

Outcome

Birth weight 3.47kg
Apgar 6 & 9
Cord gases pH 7.17 / 7.21 BD(ecf) 2 / 4



pH
 Cx Dilatation (cm)

Case 25 2156

Mrs T.T. is a 24 year old primigravida; a fit lady, and a nonsmoker. Her antenatal progress was uneventful until 36 weeks, when she developed mild hypertension without proteinuria. By 39 weeks her BP was 140/100 and she had 3+ proteinuria. She remained asymptomatic, reflexes were just slightly brisk and both haematology and biochemistry results were within normal limits. Fetal size was judged to be appropriate for dates. Labour was induced at 39 weeks using vaginal prostaglandins. Presentation cephalic 4/5; cervix closed, 50% effaced, soft. Prostin gel 2mg was given at 10.00hrs on 6.10.91, and a further 2mg dose at 16.10hrs (Bishop's score being unchanged).

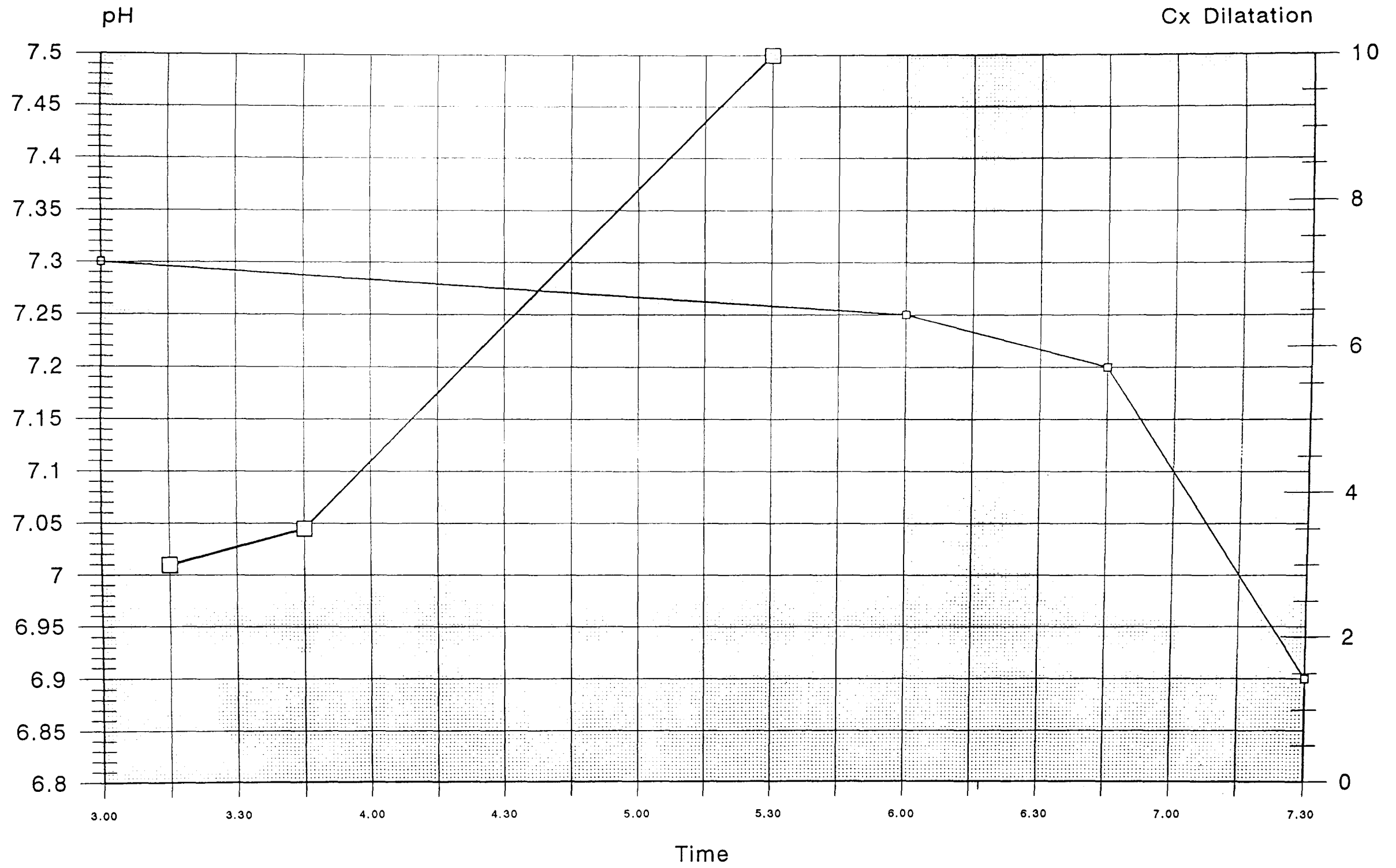
Labour events

Note; BP 120/60 to 140/85 throughout labour.

- 03.10 SR0M; liquor faintly stained with meconium.
VE Cx 3cm dilated, thin. Station -1.
- 03.40 VE Cx 3-4cm dilated. FSE applied.
- 04.25 Epidural begun.
- 04.40 Epidural main dose.
- 05.30 VE Cx fully dilated. Station 0. OA position. "Involuntary pushing".
- 06.00 Active pushing begun.
- 06.20 Maternal oxygen. Position changed.
- 07.29 Episiotomy.
- 07.30 Normal delivery of male infant.

Outcome

Birth weight 3.00kg
Apgar 6 & 7 (9 at 7 minutes)
Cord gases pH 6.87 / 6.96 BD(ecf) 17 / 18



○ pH □ Cx Dilatation (cm)

Case 26 205

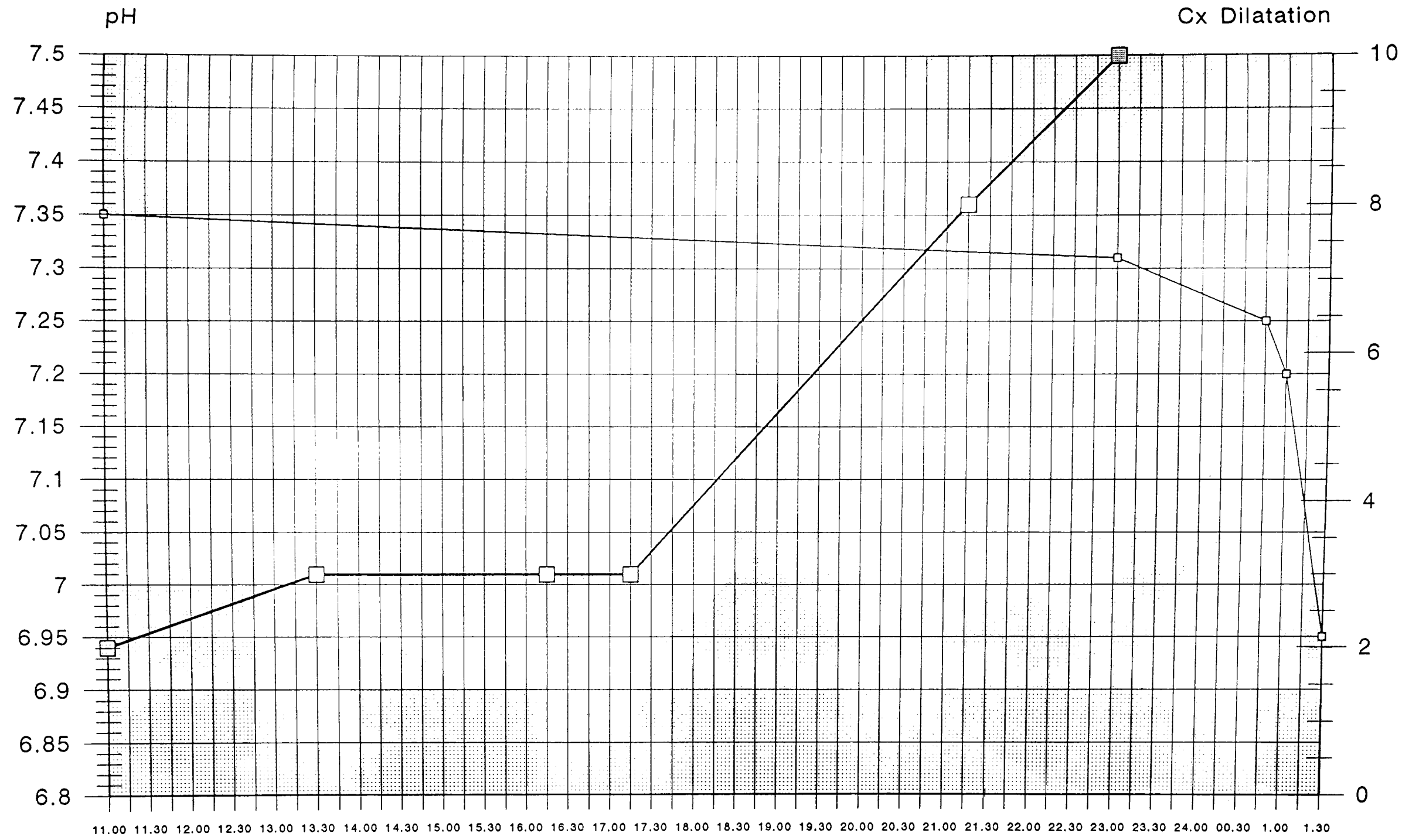
Mrs C.H. is a 28 year old primigravid lady. She is fit and does not smoke. There were no antenatal problems, and she laboured spontaneously at 41 weeks gestation. She was admitted on 7.8.90 with irregular contractions; at this stage the cervix was closed and she rested overnight on the antenatal ward. By 11.00hrs the following day, contractions were more regular and painful, and pethidine 100mg was given im.

Labour events

13.30 VE Cx 3cm dilated, 1cm thick. Station -2.
17.15 VE No change. ARM ; clear liquor. FSE applied.
17.30 Pethidine 100mg Stemetil 12.5mg im.
19.20 Epidural inserted.
19.35 Syntocinon started.
21.20 VE Cx 8cm dilated. Station 0. Position ; unsure.
22.10 Top up.
23.00 VE Cx fully dilated. Station +2. LOT position.
FBS ; pH 7.31 BE -1.7
00.45 Pushing begun.
01.10 Squatting position for 10 minutes.
01.24 Normal delivery of male infant.

Outcome

Birth weight 3.66kg
Apgar 6 & 9
Cord gases pH 6.95 / 7.13 BD(ecf) 7 / 4



□ pH ■ Cx Dilatation (cm)

Agreement results matrix for case 26

		REVIEWER																																					
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	I1	I2	J1	J2	K1	K2	L1	L2	M1	M2	N1	N2	O1	O2	P1	P2	Q1	Q2	S1	S2	R1	R2
REVIEWER	A1	100	89	66	65	70	74	73	78	86	80	75	73	78	64	74	69	90	81	75	77	87	79	69	90	81	86	79	81	66	89	82	76	86	35	72	71	61	84
	A2	89	100	72	71	76	82	79	85	77	88	81	79	85	70	81	75	99	90	67	85	83	86	77	81	88	94	87	90	58	78	90	85	95	22	76	76	64	92
	B1	66	72	100	99	59	64	78	76	72	79	79	78	76	89	77	79	71	77	61	74	69	72	72	70	71	73	73	77	83	71	74	80	68	25	90	90	65	73
	B2	65	71	99	100	57	63	77	74	72	78	77	77	74	88	77	78	71	76	61	73	68	71	71	70	70	71	74	76	82	70	74	80	69	26	89	89	65	73
	C1	70	76	59	57	100	91	70	75	63	71	73	72	75	55	73	65	75	68	77	68	84	78	80	64	77	80	67	68	54	63	78	70	71	36	66	66	53	76
	C2	74	82	64	63	91	100	73	77	75	76	75	73	77	63	84	68	81	74	77	73	74	79	89	75	80	83	79	74	66	75	81	75	77	36	69	69	56	80
	D1	73	79	78	77	70	73	100	94	87	94	98	99	94	72	95	89	78	89	53	77	79	88	63	81	89	86	85	89	70	85	89	96	74	24	86	86	61	87
	D2	78	85	76	74	75	77	94	100	85	96	96	95	100	77	94	87	84	94	58	75	76	83	67	86	85	89	90	94	72	85	84	94	81	25	84	85	60	84
	E1	86	77	72	72	63	75	87	85	100	88	86	86	85	72	87	79	77	87	69	71	76	76	68	95	77	78	86	87	77	97	79	87	74	40	77	76	56	77
	E2	80	88	79	78	71	76	94	96	88	100	95	93	96	79	94	88	87	97	61	78	78	81	69	88	84	88	95	97	73	88	85	96	85	28	84	84	61	85
	F1	75	81	79	77	73	75	98	96	86	95	100	99	96	73	96	90	80	90	55	78	80	88	64	83	88	87	86	90	70	84	88	97	76	24	86	86	62	86
	F2	73	79	78	77	72	73	99	95	86	93	99	100	95	73	96	89	78	89	55	78	80	89	63	82	87	85	85	89	70	84	89	96	74	24	86	86	61	85
	G1	78	85	76	74	75	77	94	100	85	96	96	95	100	77	94	87	84	94	58	75	76	83	67	86	85	89	90	94	72	85	84	94	81	25	84	85	60	84
	G2	64	70	89	88	55	63	72	77	72	79	73	73	77	100	74	69	69	81	52	60	59	60	61	75	61	66	79	81	82	72	63	76	67	35	94	93	48	64
	H1	74	81	77	77	73	84	95	94	87	94	96	96	94	74	100	89	81	90	59	77	79	85	73	85	83	85	92	90	73	85	87	95	80	29	83	84	60	85
	H2	69	75	79	78	65	68	89	87	79	88	90	89	87	69	89	100	74	84	53	90	74	78	60	77	78	78	82	84	75	78	80	90	72	24	79	79	68	79
	I1	90	99	71	71	75	81	78	84	77	87	80	78	84	69	81	74	100	89	67	85	82	86	76	81	88	94	87	89	59	78	90	84	96	23	77	76	65	92
	I2	81	90	77	76	68	74	89	94	87	97	90	89	94	81	90	84	89	100	63	77	75	76	70	90	78	85	96	100	73	87	81	94	87	28	81	81	59	83
	J1	75	67	61	61	77	77	53	58	69	61	55	55	58	52	59	53	67	63	100	59	75	58	87	72	57	63	64	63	67	70	61	59	67	49	53	53	52	63
	J2	77	85	74	73	68	73	77	75	71	78	78	78	75	60	77	90	85	77	59	100	81	85	67	70	83	85	74	77	61	71	87	81	81	20	73	73	72	86
	K1	87	83	69	68	84	74	79	76	76	78	80	80	76	59	79	74	82	75	75	81	100	87	65	79	85	85	72	75	64	81	89	81	78	36	73	73	61	86
	K2	79	86	72	71	78	79	88	83	76	81	88	89	83	60	85	78	86	76	58	85	87	100	68	71	97	93	73	76	52	73	98	84	81	18	79	80	66	93
	L1	69	77	72	71	80	89	63	67	68	69	64	63	67	61	73	60	76	70	87	67	65	68	100	72	68	73	75	70	70	69	71	67	73	36	62	62	62	72
	L2	90	81	70	70	64	75	81	86	95	88	83	82	86	75	85	77	81	90	72	70	79	71	72	100	72	77	89	90	81	97	74	85	78	43	75	74	55	75
	M1	81	88	71	70	77	80	89	85	77	84	88	87	85	61	83	78	88	78	57	83	85	97	68	72	100	95	75	78	54	76	96	84	83	19	79	80	66	95
	M2	86	94	73	71	80	83	86	89	78	88	87	85	89	66	85	78	94	85	63	85	85	93	73	77	95	100	82	85	56	77	94	85	90	20	80	80	65	95
	N1	79	87	73	74	67	79	85	90	86	95	86	85	90	79	92	82	87	96	64	74	72	73	75	89	75	82	100	96	74	86	78	90	89	33	79	78	58	82
	N2	81	90	77	76	68	74	89	94	87	97	90	89	94	81	90	84	89	100	63	77	75	76	70	90	78	85	96	100	73	87	81	94	87	28	81	81	59	83
	O1	66	58	83	82	54	66	70	72	77	73	70	70	72	82	73	75	59	73	67	61	64	52	70	81	54	56	74	73	100	82	55	72	56	52	77	77	55	57
	O2	89	78	71	70	63	75	85	85	97	88	84	84	85	72	85	78	78	87	70	71	81	73	69	97	76	77	86	87	82	100	76	85	75	43	77	76	56	76
	P1	82	90	74	74	78	81	89	84	79	85	88	89	84	63	87	80	90	81	61	87	89	98	71	74	96	94	78	81	55	76	100	88	86	19	80	80	65	96
	P2	76	85	80	80	70	75	96	94	87	96	97	96	94	76	95	90	84	94	59	81	81	84	67	85	84	85	90	94	72	85	88	100	80	26	85	84	60	87
	Q1	86	95	68	69	71	77	74	81	74	85	76	74	81	67	80	72	96	87	67	81	78	81	73	78	83	90	89	87	56	75	86	80	100	24	73	73	63	91
	Q2	35	22	25	26	36	36	24	25	40	28	24	24	25	35	29	24	23	28	49	20	36	18	36	43	19	20	33	28	52	43	19	26	24	100	26	26	18	22
S1	72	76	90	89	66	69	86	84	77	84	86	86	84	94	83	79	77	81	53	73	73	79	62	75	79	80	79	81	77	77	80	85	73	26	100	100	58	79	
S2	71	76	90	89	66	69	86	85	76	84	86	86	85	93	84	79	76	81	53	73	73	80	62	74	80	80	78	81	77	76	80	84	73	26	100	100	58	80	
R1	61	64	65	65	53	56	61	60	56	61	62	61	60	48	60	68	65	59	52	72	61	66	62	55	66	65	58	59	55	56	65	60	63	18	58	58	100	67	
R2	84	92	73	73	76	80	87	84	77	85	86	85	84	64	85	79	92	83	63	86	86	93	72	75	95	95	82	83	57	76	96	87	91	22	79	80	67	100	

Case 27 1288

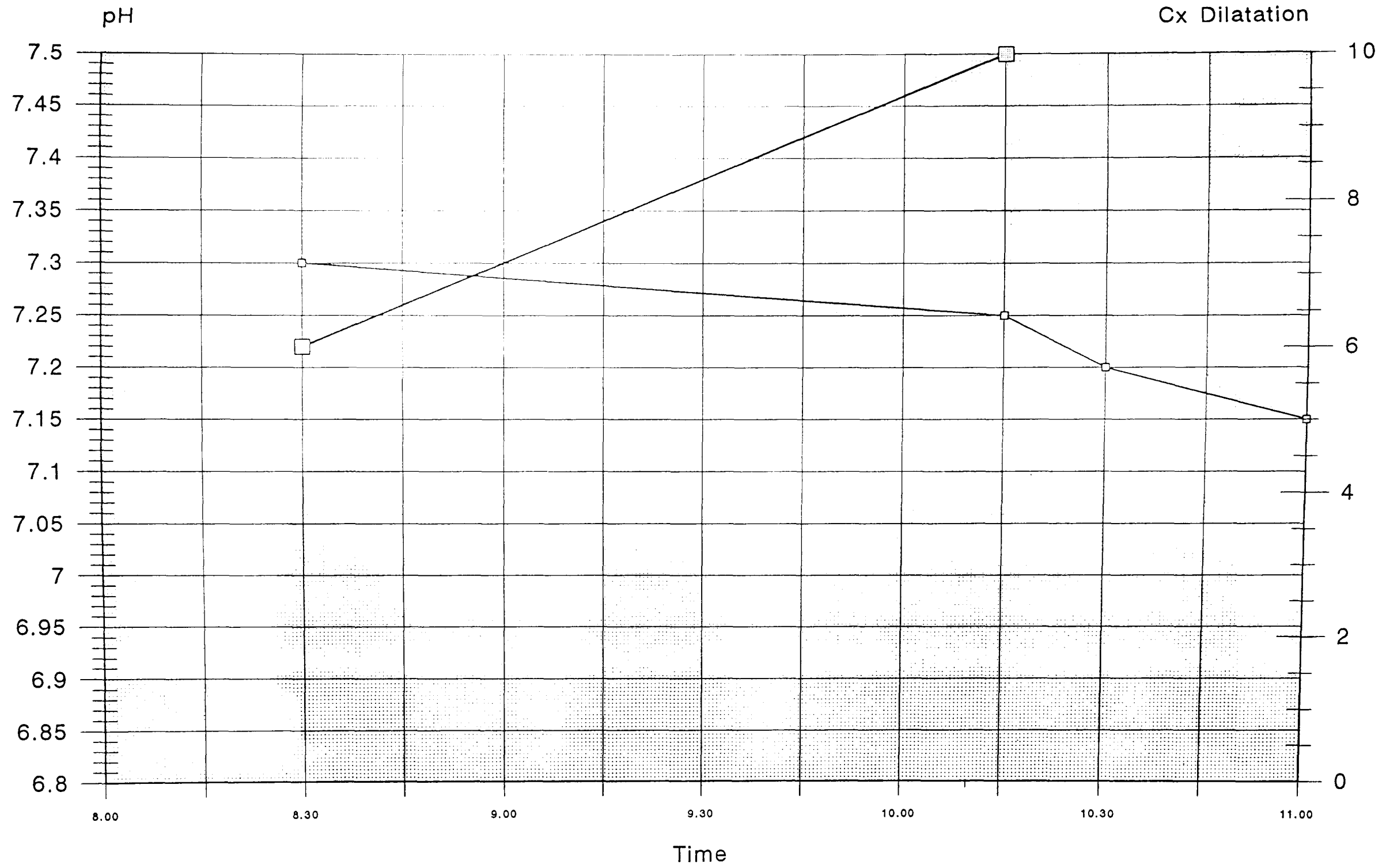
Miss D.F. is a fit 27 year old primigravida. She smokes 5 cigarettes per day. Her pregnancy was quite straightforward and she was admitted in spontaneous labour at 39 weeks on 4/4/91 at 08.30hrs. The uterus was term size. Cephalic 1/5. VE (at 09.30hrs) Cx 6cm dilated, 0.5cm thick. Station -1. (Membranes had already ruptured; liquor clear.)

Labour events

09.45 Pethidine 100mg, stemetil 12.5mg im.
11.15 VE Cx fully dilated. Station +1.
11.30 Pushing begun.
11.50 Episiotomy.
12.00 Normal delivery of female infant.

Outcome

Birth weight 3.14kg
Apgar 9 & 9
Cord gases pH 7.24 BD(ecf) - / 5



□ pH ○ Cx Dilatation (cm)

Agreement results matrix for case 27

		REVIEWER																																					
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	I1	I2	J1	J2	K1	K2	L1	L2	M1	M2	N1	N2	O1	O2	P1	P2	Q1	Q2	S1	S2	R1	R2
REVIEWER	A1	100	52	58	85	97	93	98	100	85	95	97	99	100	77	100	95	67	89	84	91	99	95	97	99	52	94	92	89	93	94	54	95	94	47	76	77	87	76
	A2	52	100	70	48	50	51	52	52	48	54	52	51	52	90	52	54	90	53	47	50	51	54	52	51	20	54	48	51	51	55	95	54	54	80	90	90	43	44
	B1	58	70	100	72	57	71	69	58	72	47	69	58	58	81	58	47	71	69	71	71	58	47	69	58	13	47	57	45	71	67	74	47	47	95	80	81	60	62
	B2	85	48	72	100	84	97	92	85	100	75	92	85	85	68	85	75	73	94	99	95	85	75	92	85	37	74	80	70	97	90	50	75	74	54	67	68	83	87
	C1	97	50	57	84	100	91	96	97	84	93	97	99	97	75	97	93	66	87	81	89	99	95	97	99	57	90	97	93	91	92	56	95	90	46	76	75	87	77
	C2	93	51	71	97	91	100	97	93	97	82	96	92	93	72	93	82	76	97	96	99	92	82	96	92	38	81	88	78	100	94	53	82	81	55	71	72	88	80
	D1	98	52	69	92	96	97	100	98	92	89	100	97	98	75	98	89	73	93	91	96	97	88	100	97	37	88	93	84	97	96	55	88	88	54	74	75	90	78
	D2	100	52	58	85	97	93	98	100	85	95	97	99	100	77	100	95	67	89	84	91	99	95	97	99	52	94	92	89	93	94	54	95	94	47	76	77	87	76
	E1	85	48	72	100	84	97	92	85	100	75	92	85	85	68	85	75	73	94	99	95	85	75	92	85	37	74	80	70	97	90	50	75	74	54	67	68	83	87
	E2	95	54	47	75	93	82	89	95	75	100	88	95	95	79	95	100	71	86	74	80	95	99	88	95	57	99	87	94	82	93	55	99	99	42	78	79	78	73
	F1	97	52	69	92	97	96	100	97	92	88	100	98	97	74	97	88	72	93	90	96	98	89	100	98	39	86	94	85	96	96	56	89	86	53	75	74	89	79
	F2	99	51	58	85	99	92	97	99	85	95	98	100	99	76	99	95	67	89	83	91	100	95	98	100	55	92	94	91	92	94	55	95	92	47	77	76	86	77
	G1	100	52	58	85	97	93	98	100	85	95	97	99	100	77	100	95	67	89	84	91	99	95	97	99	52	94	92	89	93	94	54	95	94	47	76	77	87	76
	G2	77	90	81	68	75	72	75	77	68	79	74	76	77	100	77	79	88	74	67	70	76	78	74	76	32	78	72	74	72	77	90	78	78	84	100	100	63	63
	H1	100	52	58	85	97	93	98	100	85	95	97	99	100	77	100	95	67	89	84	91	99	95	97	99	52	94	92	89	93	94	54	95	94	47	76	77	87	76
	H2	95	54	47	75	93	82	89	95	75	100	88	95	95	79	95	100	71	86	74	80	95	99	88	95	57	99	87	94	82	93	55	99	99	42	78	79	78	73
	I1	67	90	71	73	66	76	73	67	73	71	72	67	67	88	67	71	100	83	73	74	67	71	72	67	29	70	63	68	76	81	86	71	70	72	88	88	65	61
	I2	89	53	69	94	87	97	93	89	94	86	93	89	89	74	89	86	83	100	93	95	89	86	93	89	42	85	84	82	97	97	54	86	85	56	73	74	86	78
	J1	84	47	71	99	81	96	91	84	99	74	90	83	84	67	84	74	73	93	100	94	83	72	90	83	34	75	78	67	96	89	48	72	75	55	66	67	84	86
	J2	91	50	71	95	89	99	96	91	95	80	96	91	91	70	91	80	74	95	94	100	91	79	96	91	41	79	90	79	99	92	54	79	79	55	70	70	89	79
	K1	99	51	58	85	99	92	97	99	85	95	98	100	99	76	99	95	67	89	83	91	100	95	98	100	55	92	94	91	92	94	55	95	92	47	77	76	86	77
	K2	95	54	47	75	95	82	88	95	75	99	89	95	95	78	95	99	71	86	72	79	95	100	89	95	59	97	89	96	82	93	56	100	97	41	79	78	77	74
	L1	97	52	69	92	97	96	100	97	92	88	100	98	97	74	97	88	72	93	90	96	98	89	100	98	39	86	94	85	96	96	56	89	86	53	75	74	89	79
	L2	99	51	58	85	99	92	97	99	85	95	98	100	99	76	99	95	67	89	83	91	100	95	98	100	55	92	94	91	92	94	55	95	92	47	77	76	86	77
	M1	52	20	13	37	57	38	37	52	37	57	39	55	52	32	52	57	29	42	34	41	55	59	39	55	100	54	61	67	38	43	36	59	54	9	34	32	64	40
	M2	94	54	47	74	90	81	88	94	74	99	86	92	94	78	94	99	70	85	75	79	92	97	86	92	54	100	85	91	81	92	53	97	100	42	77	78	79	72
	N1	92	48	57	80	97	88	93	92	80	87	94	94	92	72	92	87	63	84	78	90	94	89	94	94	61	85	100	95	88	89	58	89	85	46	73	72	88	74
	N2	89	51	45	70	93	78	84	89	70	94	85	91	89	74	89	94	68	82	67	79	91	96	85	91	67	91	95	100	78	88	59	96	91	40	76	74	79	71
	O1	93	51	71	97	91	100	97	93	97	82	96	92	93	72	93	82	76	97	96	99	92	82	96	92	38	81	88	78	100	94	53	82	81	55	71	72	88	80
	O2	94	55	67	90	92	94	96	94	90	93	96	94	94	77	94	93	81	97	89	92	94	93	96	94	43	92	89	88	94	100	56	93	92	55	76	77	88	76
	P1	54	95	74	50	56	53	55	54	50	55	56	55	54	90	54	55	86	54	48	54	55	56	56	55	36	53	58	59	53	56	100	56	53	80	91	90	54	47
	P2	95	54	47	75	95	82	88	95	75	99	89	95	95	78	95	99	71	86	72	79	95	100	89	95	59	97	89	96	82	93	56	100	97	41	79	78	77	74
	Q1	94	54	47	74	90	81	88	94	74	99	86	92	94	78	94	99	70	85	75	79	92	97	86	92	54	100	85	91	81	92	53	97	100	42	77	78	79	72
	Q2	47	80	95	54	46	55	54	47	54	42	53	47	47	84	47	42	72	56	55	55	47	41	53	47	9	42	46	40	55	55	80	41	42	100	83	84	46	48
S1	76	90	80	67	76	71	74	76	67	78	75	77	76	100	76	78	88	73	66	70	77	79	75	77	34	77	73	76	71	76	91	79	77	83	100	100	62	64	
S2	77	90	81	68	75	72	75	77	68	79	74	76	77	100	77	79	88	74	67	70	76	78	74	76	32	78	72	74	72	77	90	78	78	84	100	100	63	63	
R1	87	43	60	83	87	88	90	87	83	78	89	86	87	63	87	78	65	86	84	89	86	77	89	86	64	79	88	79	88	88	54	77	79	46	62	63	100	69	
R2	76	44	62	87	77	80	78	76	87	73	79	77	76	63	76	73	61	78	86	79	77	74	79	77	40	72	74	71	80	76	47	74	72	48	64	63	69	100	

Case 28 1232

Miss E.M. is a 20 year old primigravida. She is generally fit, but smokes 3 cigarettes per day. There were no antenatal problems and she laboured spontaneously at 40 weeks gestation. When she arrived on labour ward at 19.15hrs on 18.3.91, presentation was cephalic 2/5; cervix closed and uneffaced; station -2.

Labour events

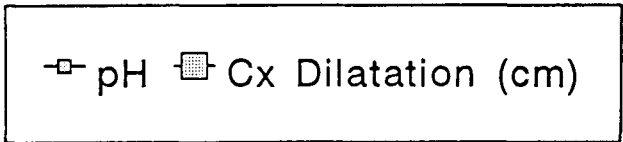
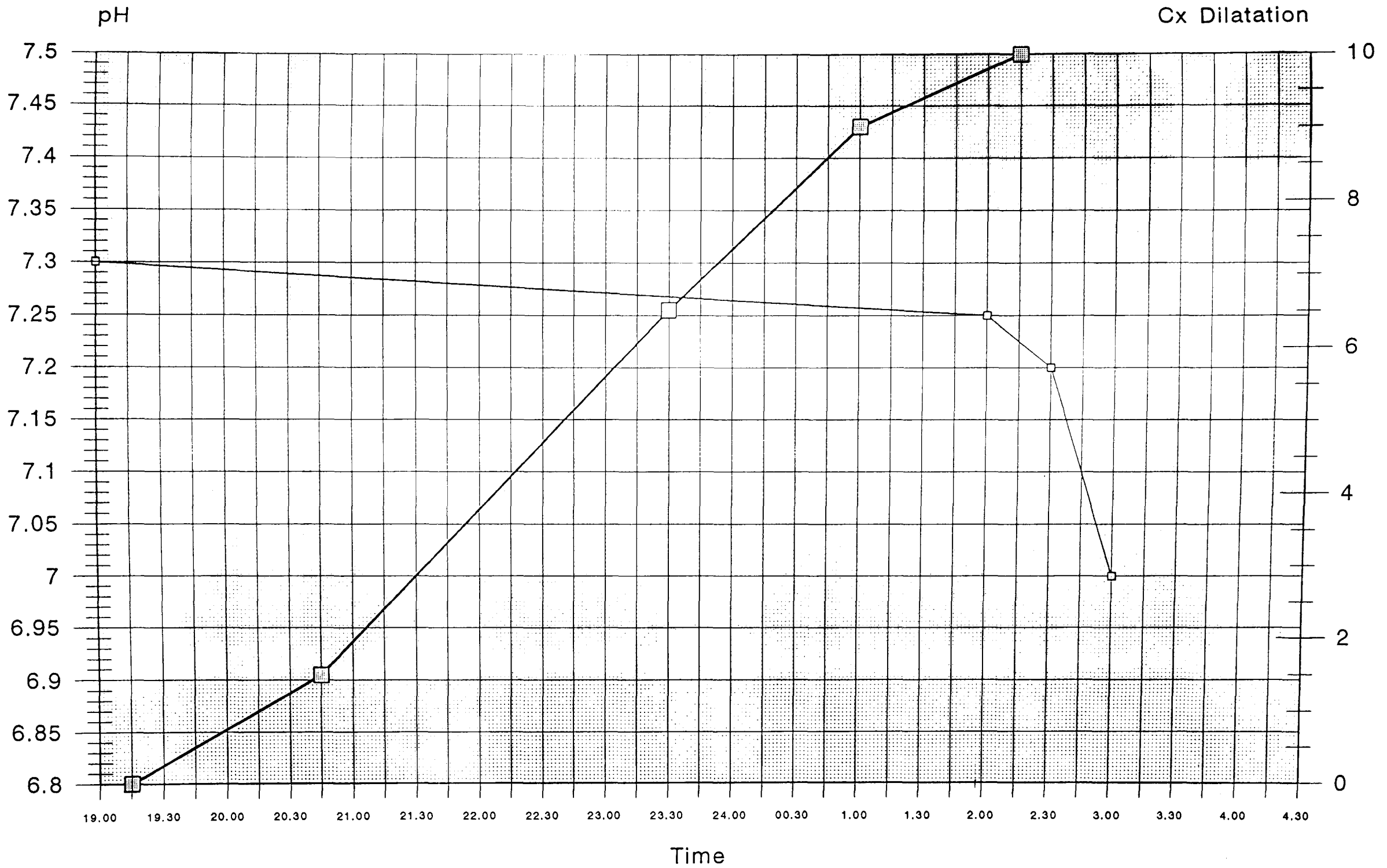
- 20.05 Pethidine 100mg, Phenergan 25mg given.
- 20.40 SR0M; liquor stained with thick meconium. VE Cx 1-2cm dilated, thick. Station -2. FSE applied.
- 21.40 Epidural begun.
- 22.30 Epidural complete.
- 23.30 VE Cx 6-7cm dilated, thin. Station -1.
- 01.00 VE Cx 9cm dilated. Station -1.
- 01.10 Top up.
- 02.10 VE Cx fully dilated. Pushing begun.
- 03.06 Male infant delivered by Ventouse extraction (after attempted manual rotation).

Outcome

Birth weight 3.15kg

Apgar 5 & 9

Cord gases pH 6.97 / 7.10 BD(ecf) 7 / 6



Agreement results matrix for case 28

		REVIEWER																																					
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	I1	I2	J1	J2	K1	K2	L1	L2	M1	M2	N1	N2	O1	O2	P1	P2	Q1	Q2	S1	S2	R1	R2
REVIEWER	A1	100	94	47	81	0	71	61	50	62	64	53	60	65	66	61	61	66	62	56	65	60	64	54	62	45	63	67	65	66	66	82	62	64	40	86	86	28	56
	A2	94	100	49	91	0	77	69	49	70	69	59	69	72	73	58	69	75	69	61	72	69	75	54	68	59	75	64	73	75	75	87	70	72	42	94	94	34	61
	B1	47	49	100	57	25	63	78	67	78	59	89	78	55	56	73	65	60	57	67	59	49	56	58	60	47	57	49	59	60	56	60	66	53	88	54	54	67	88
	B2	81	91	57	100	0	84	80	56	80	80	68	80	81	81	67	78	88	78	66	83	74	84	62	80	65	84	80	86	86	82	85	78	77	43	95	95	34	70
	C1	0	0	25	0	100	16	12	11	12	0	33	12	0	0	15	0	0	0	16	0	0	0	18	0	0	0	0	0	0	0	16	19	0	47	0	0	34	34
	C2	71	77	63	84	16	100	77	69	77	66	68	77	66	66	83	79	73	72	83	71	62	70	67	69	54	70	71	72	71	69	88	81	65	51	80	80	51	67
	D1	61	69	78	80	12	77	100	68	97	77	88	99	80	80	78	76	89	78	79	81	80	88	70	81	75	88	54	84	87	83	79	89	77	61	79	79	50	81
	D2	50	49	67	56	11	69	68	100	74	71	65	70	64	67	86	74	58	56	85	66	39	51	81	66	44	52	63	64	60	54	58	68	61	56	49	49	60	69
	E1	62	70	78	80	12	77	97	74	100	82	87	98	83	84	81	80	88	79	85	84	75	85	76	81	69	85	58	87	87	84	82	92	83	65	77	77	50	82
	E2	64	69	59	80	0	66	77	71	82	100	63	78	94	96	76	83	88	83	79	92	68	81	77	93	68	81	74	91	90	85	71	82	91	47	74	74	38	79
	F1	53	59	89	68	33	68	88	65	87	63	100	88	65	66	73	70	74	64	69	69	67	73	60	67	61	72	47	71	71	70	70	76	64	86	65	65	64	92
	F2	60	69	78	80	12	77	99	70	98	78	88	100	78	79	78	78	88	78	81	82	77	88	71	79	72	88	56	85	85	81	80	89	79	62	78	78	50	81
	G1	65	72	55	81	0	66	80	64	83	94	65	78	100	98	70	80	91	80	78	89	80	86	72	92	74	85	64	89	93	92	74	83	93	49	80	80	38	77
	G2	66	73	56	81	0	66	80	67	84	96	66	79	98	100	72	83	91	82	80	90	77	86	73	92	71	86	66	90	93	91	75	85	96	51	78	78	38	78
	H1	61	58	73	67	15	83	78	86	81	76	73	78	70	72	100	82	71	66	88	78	53	62	79	73	52	61	68	77	71	67	73	83	64	57	60	60	54	74
	H2	61	69	65	78	0	79	76	74	80	83	70	78	80	83	82	100	85	81	85	83	72	82	62	80	66	83	59	84	86	82	72	80	83	53	75	75	55	68
	I1	66	75	60	88	0	73	89	58	88	88	74	88	91	91	71	85	100	86	72	91	88	96	67	91	83	95	60	96	98	94	77	87	87	45	85	85	40	76
	I2	62	69	57	78	0	72	78	56	79	83	64	78	80	82	66	81	86	100	74	83	74	84	74	81	68	84	58	84	86	84	71	79	82	44	77	77	48	69
	J1	56	61	67	66	16	83	79	85	85	79	69	81	78	80	88	85	72	74	100	73	59	69	76	73	53	70	58	72	73	72	77	84	80	63	65	65	56	74
	J2	65	72	59	83	0	71	81	66	84	92	69	82	89	90	78	83	91	83	73	100	76	85	74	90	77	85	69	96	93	90	73	85	86	44	78	78	38	77
	K1	60	69	49	74	0	62	80	39	75	68	67	77	80	77	53	72	88	74	59	76	100	94	47	77	87	93	39	77	87	92	72	74	80	42	84	84	40	63
	K2	64	75	56	84	0	70	88	51	85	81	73	88	86	86	62	82	96	84	69	85	94	100	58	85	84	100	51	89	93	93	76	83	88	45	86	86	41	71
	L1	54	54	58	62	18	67	70	81	76	77	60	71	72	73	79	62	67	74	76	74	47	58	100	75	53	58	66	73	67	62	69	81	67	49	54	54	51	72
	L2	62	68	60	80	0	69	81	66	81	93	67	79	92	92	73	80	91	81	73	90	77	85	75	100	77	84	70	90	92	84	70	80	85	44	75	75	39	79
	M1	45	59	47	65	0	54	75	44	69	68	61	72	74	71	52	66	83	68	53	77	87	84	53	77	100	84	38	78	82	83	57	68	70	34	68	68	44	62
	M2	63	75	57	84	0	70	88	52	85	81	72	88	85	86	61	83	95	84	70	85	93	100	58	84	84	100	51	88	93	93	76	82	88	46	86	86	42	71
	N1	67	64	49	80	0	71	54	63	58	74	47	56	64	66	68	59	60	58	58	69	39	51	66	70	38	51	100	68	61	53	64	58	60	36	65	65	26	62
	N2	65	73	59	86	0	72	84	64	87	91	71	85	89	90	77	84	96	84	72	96	77	89	73	90	78	88	68	100	95	89	73	86	84	43	79	79	38	77
	O1	66	75	60	86	0	71	87	60	87	90	71	85	93	93	71	86	98	86	73	93	87	93	67	92	82	93	61	95	100	96	77	87	89	46	86	86	40	77
	O2	66	75	56	82	0	69	83	54	84	85	70	81	92	91	67	82	94	84	72	90	92	93	62	84	83	93	53	89	96	100	79	85	91	48	85	85	40	72
	P1	82	87	60	85	16	88	79	58	82	71	70	80	74	75	73	72	77	71	77	73	72	76	69	70	57	76	64	73	77	79	100	89	75	56	88	88	40	68
	P2	62	70	66	78	19	81	89	68	92	82	76	89	83	85	83	80	87	79	84	85	74	83	81	80	68	82	58	86	87	85	89	100	82	57	75	75	45	78
Q1	64	72	53	77	0	65	77	61	83	91	64	79	93	96	64	83	87	82	80	86	80	88	67	85	70	88	60	84	89	91	75	82	100	53	79	79	40	72	
Q2	40	42	88	43	47	51	61	56	65	47	86	62	49	51	57	53	45	44	63	44	42	45	49	44	34	46	36	43	46	48	56	57	53	100	44	44	73	89	
S1	86	94	54	95	0	80	79	49	77	74	65	78	80	78	60	75	85	77	65	78	84	86	54	75	68	86	65	79	86	85	88	75	79	44	100	100	35	65	
S2	86	94	54	95	0	80	79	49	77	74	65	78	80	78	60	75	85	77	65	78	84	86	54	75	68	86	65	79	86	85	88	75	79	44	100	100	35	65	
R1	28	34	67	34	34	51	50	60	50	38	64	50	38	38	54	55	40	48	56	38	40	41	51	39	44	42	26	38	40	40	40	45	40	73	35	35	100	57	
R2	56	61	88	70	34	67	81	69	82	79	92	81	77	78	74	68	76	69	74	77	63	71	72	79	62	71	62	77	77	72	68	78	72	89	65	65	57	100	

Case 29 194

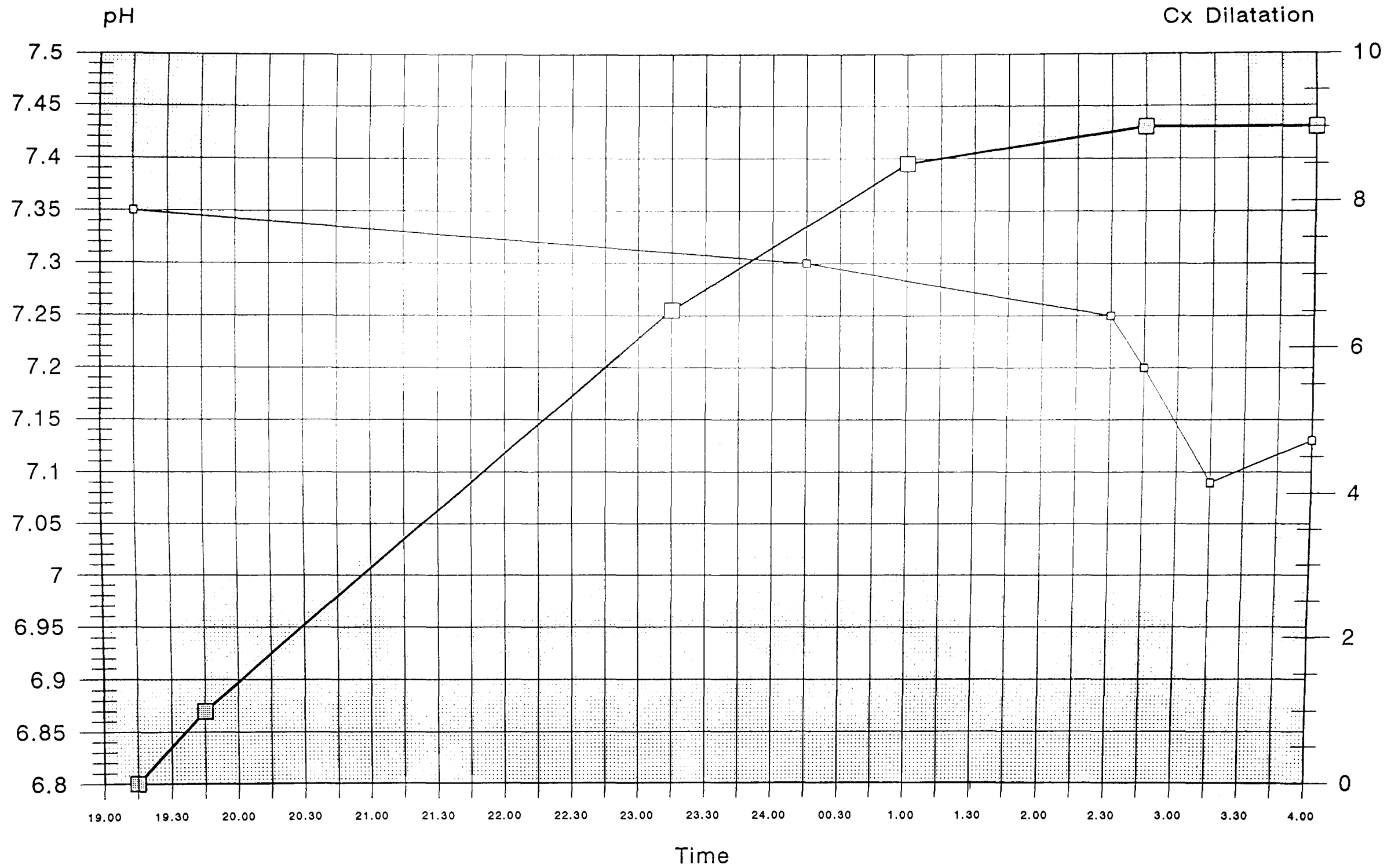
Mrs S.D. is a 28 year old primigravida; fit and well and a nonsmoker. Her pregnancy was straightforward. She was admitted at 41 weeks with spontaneous rupture of the membranes; this had probably occurred 4 days previously. The uterus was felt to be term size; cephalic 4/5. On VE clear liquor was seen and the cervix closed and uneffaced. Since she had no contractions labour was induced with IV syntocinon, which was commenced at 15.20hrs on 11.8.90.

Labour events

- 19.15 VE Cx closed, uneffaced. Station -3.
- 19.45 VE Cx 1cm dilated, 0.5cm thick. Forewaters ruptured; clear liquor.
- 21.20 Epidural inserted.
- 23.20 VE Cx 6-7cm dilated, thick. Station -1. FSE applied.
- 23.40 Top up.
- 00.15 FBS; pH 7.30 BE -7.
- 01.05 VE Cx 8-9cm dilated, thin. Station -1.
- 01.40 Top up.
- 02.45 VE Cx 9cm dilated. Station -1.
Maternal oxygen given; position changed.
- 03.10 FBS; pH 7.09 (no BE) Thick meconium now present.
Decision for C/S.
- 03.35 Top up.
- 03.50 Top up.
- 04.00 Male infant delivered by emergency C/S under epidural. Indication; low scalp pH; immediate vaginal delivery not possible. Fresh meconium noted at operation.

Outcome

Birth weight 3.21kg
Apgar 6 & 6 (10 at 10 minutes)
Cord gases pH 7.13 / 7.20 BD(ecf) 4 / 6



□ pH ■ Cx Dilatation (cm)

Agreement results matrix for case 29

		REVIEWER																																					
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	I1	I2	J1	J2	K1	K2	L1	L2	M1	M2	N1	N2	O1	O2	P1	P2	Q1	Q2	S1	S2	R1	R2
REVIEWER	A1	100	83	78	74	89	89	90	92	94	91	93	89	92	83	82	89	88	90	88	89	92	84	89	96	96	97	77	90	89	89	91	93	64	53	76	76	63	51
	A2	83	100	71	78	83	86	70	80	79	78	86	75	83	83	85	85	89	85	89	81	86	86	81	81	89	79	77	86	83	84	88	84	70	68	79	79	64	43
	B1	78	71	100	90	87	89	82	86	85	87	66	76	78	68	92	78	78	78	80	91	85	54	88	82	68	74	94	89	88	77	73	65	83	73	70	70	66	56
	B2	74	78	90	100	85	87	74	84	80	83	67	75	76	79	90	79	80	79	81	90	83	59	86	79	69	71	90	87	86	79	77	66	91	89	81	81	63	50
	C1	89	83	87	85	100	94	86	95	92	95	84	88	89	82	90	94	91	92	94	93	95	76	95	92	85	85	85	93	99	92	88	83	73	66	79	78	63	51
	C2	89	86	89	87	94	100	87	92	95	93	82	89	88	84	94	86	93	90	92	93	90	72	93	93	83	87	86	95	94	89	91	84	78	69	81	81	62	48
	D1	90	70	82	74	86	87	100	89	90	89	81	88	88	71	85	80	84	82	77	87	89	70	83	86	81	91	80	87	85	83	80	82	66	54	68	67	64	57
	D2	92	80	86	84	95	92	89	100	95	98	86	91	89	83	88	96	94	96	90	96	93	79	95	96	89	89	87	95	96	94	91	86	71	63	81	81	65	53
	E1	94	79	85	80	92	95	90	95	100	98	85	93	87	85	88	90	95	94	91	93	88	76	92	97	88	92	84	93	93	93	93	89	72	60	79	79	59	53
	E2	91	78	87	83	95	93	89	98	98	100	84	89	87	83	88	94	93	94	91	96	91	75	95	95	87	88	87	95	95	93	90	86	72	64	78	78	63	54
	F1	93	86	66	67	84	82	81	86	85	84	100	90	87	82	73	89	87	89	86	82	86	96	83	90	96	91	69	82	82	91	91	92	57	50	78	78	61	53
	F2	89	75	76	75	88	89	88	91	93	89	90	100	90	85	83	88	91	89	84	88	84	82	87	94	88	93	76	89	87	90	94	89	66	55	81	81	55	52
	G1	92	83	78	76	89	88	88	89	87	87	87	90	100	79	85	88	85	83	86	85	94	81	88	89	90	96	76	91	90	82	88	92	65	58	73	72	57	53
	G2	83	83	68	79	82	84	71	83	85	83	82	85	79	100	79	86	88	86	84	82	78	78	82	87	84	83	73	83	82	85	91	85	81	72	96	96	52	43
	H1	82	85	92	90	90	94	85	88	88	88	73	83	85	79	100	83	89	83	85	95	87	64	85	86	75	81	95	95	90	83	85	74	85	78	80	80	68	50
	H2	89	85	78	79	94	86	80	96	90	94	89	88	88	86	83	100	95	97	92	92	90	87	94	93	93	87	82	92	94	96	94	89	67	63	82	82	62	53
	I1	88	89	78	80	91	93	84	94	95	93	87	91	85	88	89	95	100	97	92	92	88	84	90	92	90	88	84	92	91	96	96	90	72	66	84	84	62	51
	I2	90	85	78	79	92	90	82	96	94	94	89	89	83	86	83	97	97	100	95	92	89	87	91	94	92	86	82	89	93	99	94	89	68	61	83	83	62	53
	J1	88	89	80	81	94	92	77	90	91	91	86	84	86	84	85	92	92	95	100	88	90	83	89	91	89	83	79	90	95	94	90	86	70	64	80	80	59	50
	J2	89	81	91	90	93	93	87	96	93	96	82	88	85	82	95	92	92	92	88	100	90	73	93	93	84	86	92	95	94	91	89	82	80	71	81	81	70	52
	K1	92	86	85	83	95	90	89	93	88	91	86	84	94	78	87	90	88	89	90	90	100	78	93	89	90	89	82	91	96	88	84	86	70	63	74	74	65	52
	K2	84	86	54	59	76	72	70	79	76	75	96	82	81	78	64	87	84	87	83	73	78	100	75	81	94	84	60	73	75	87	87	91	48	45	75	74	57	54
	L1	89	81	88	86	95	93	83	95	92	95	83	87	88	82	85	94	90	91	89	93	93	75	100	93	86	86	84	94	95	90	88	83	71	65	79	79	62	50
	L2	96	81	82	79	92	93	86	96	97	95	90	94	89	87	86	93	92	94	91	93	89	81	93	100	93	93	81	93	93	93	94	90	69	57	83	83	61	49
	M1	96	89	68	69	85	83	81	89	88	87	96	88	90	84	75	93	90	92	89	84	90	94	86	93	100	94	71	85	86	91	93	96	58	51	79	78	60	52
	M2	97	79	74	71	85	87	91	89	92	88	91	93	96	83	81	87	88	86	83	86	89	84	86	93	94	100	74	87	86	85	92	96	62	51	75	74	57	54
	N1	77	77	94	90	85	86	80	87	84	87	69	76	76	73	95	82	84	82	79	92	82	60	84	81	71	74	100	90	85	81	78	68	86	82	75	74	66	57
	N2	90	86	89	87	93	95	87	95	93	95	82	89	91	83	95	92	92	89	90	95	91	73	94	93	85	87	90	100	94	88	91	83	76	70	81	81	65	48
	O1	89	83	88	86	99	94	85	96	93	95	82	87	90	82	90	94	91	93	95	94	96	75	95	93	86	86	85	94	100	92	88	83	73	66	78	78	63	51
	O2	89	84	77	79	92	89	83	94	93	93	91	90	82	85	83	96	96	99	94	91	88	87	90	93	91	85	81	88	92	100	93	88	67	61	83	83	63	54
	P1	91	88	73	77	88	91	80	91	93	90	91	94	88	91	85	94	96	94	90	89	84	87	88	94	93	92	78	91	88	93	100	94	69	61	86	86	58	48
	P2	93	84	65	66	83	84	82	86	89	86	92	89	92	85	74	89	90	89	86	82	86	91	83	90	96	96	68	83	83	88	94	100	57	50	74	74	54	55
	Q1	64	70	83	91	73	78	66	71	72	72	57	66	65	81	85	67	72	68	70	80	70	48	71	69	58	62	86	76	73	67	69	57	100	96	84	84	55	47
	Q2	53	68	73	89	66	69	54	63	60	64	50	55	58	72	78	63	66	61	64	71	63	45	65	57	51	51	82	70	66	61	61	50	96	100	75	75	51	46
S1	76	79	70	81	79	81	68	81	79	78	78	81	73	96	80	82	84	83	80	81	74	75	79	83	79	75	75	81	78	83	86	74	84	75	100	100	55	40	
S2	76	79	70	81	78	81	67	81	79	78	78	81	72	96	80	82	84	83	80	81	74	74	79	83	78	74	74	81	78	83	86	74	84	75	100	100	55	40	
R1	63	64	66	63	63	62	64	65	59	63	61	55	57	52	68	62	62	62	59	70	65	57	62	61	60	57	66	65	63	63	58	54	55	51	55	55	100	61	
R2	51	43	56	50	51	48	57	53	53	54	53	52	53	43	50	53	51	53	50	52	52	54	50	49	52	54	57	48	51	54	48	55	47	46	40	40	61	100	

Case 30 359

Miss J.B. is an 18 year old primigravida. She is a fit lady and does not smoke. Her pregnancy was straightforward until term, when a mild degree of hypertension was noted (BP 140/90). By 41 weeks BP was 140/100 and she had a trace of proteinuria. The baby was felt to be appropriately grown; presentation cephalic 4/5, and the cervix 1cm dilated, partly effaced. Labour was now induced using vaginal prostaglandins; prostin 3mg was given at 12.15hrs on 16.9.90. Contractions gradually became established over the next 18 hours.

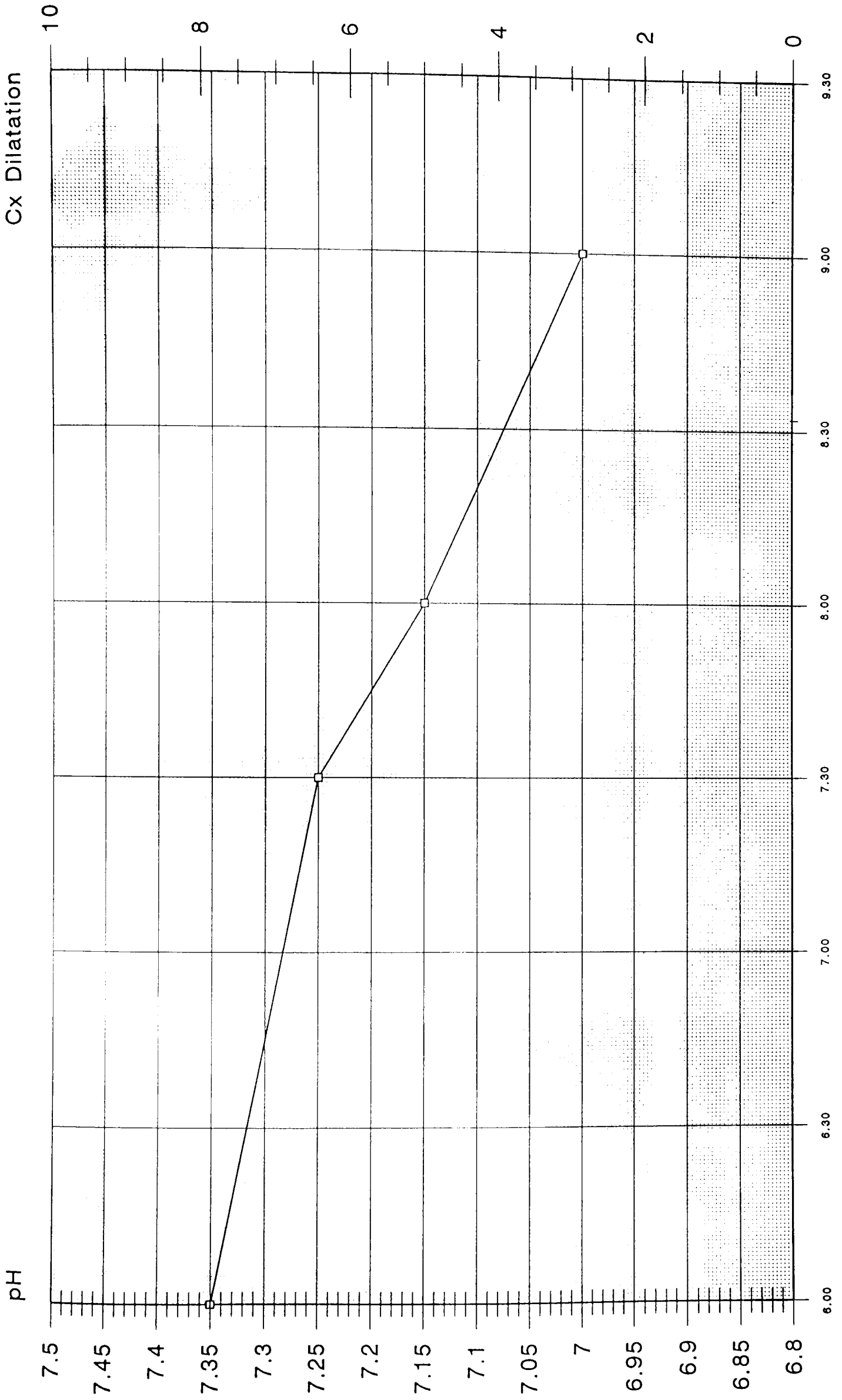
Labour events

Note; BP stable throughout labour at 130/90-100

- 06.10 Contractions now more painful. Continuous monitoring begun.
- 06.30 VE Cx fully dilated. Station 0. ARM performed; clear liquor. Position LOA.
- 07.43 FSE applied.
- 07.50 Pushing begun.
- 08.05 In kneeling position.
- 08.20 Maternal position changed, oxygen given.
- 09.10 Episiotomy.
- 09.15 Normal delivery of male infant.

Outcome

Birth weight 3.66kg
Apgar 6 & 9
Cord gases pH 6.88 / 6.97 BD(ecf) 22 / 15



□ pH

Case 31 2324

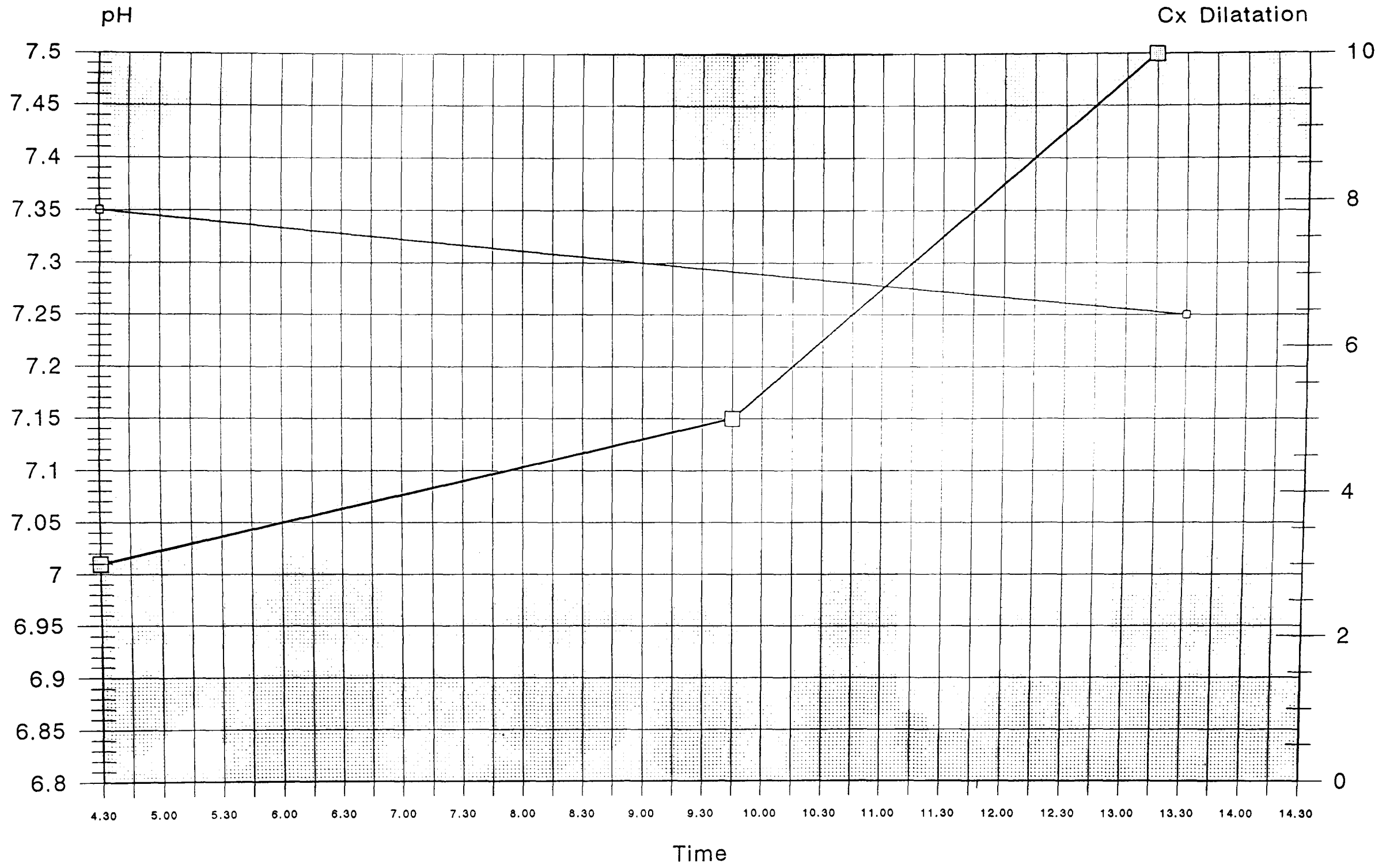
Mrs J.W. is a 24 year old lady expecting her second baby. She is a fit lady and does not smoke. Her first child was delivered vaginally as a breech. He was adopted, and there is no record of his weight. Her current pregnancy progressed well, and she was admitted with spontaneous rupture of the membranes at 39 weeks, on 20.11.91. The uterus was felt to be appropriate for dates, presentation cephalic 3/5. VE (at 04.30hrs) Cx 3cm dilated. Contractions were mild and 5 hours later (at 09.50hrs) the cervix was still just 5cm dilated. Forewater rupture was performed; clear liquor was seen.

Labour events

10.50 Epidural begun.
11.10 Epidural complete.
13.15 VE Cx fully dilated. Station +1. LOA position.
13.45 Pushing begun.
14.26 Normal delivery of female infant.

Outcome

Birth weight 3.4kg
Apgar 9 & 9
Cord gases pH 7.25 / 7.37 BD(ecf) 0 / 0



○ pH □ Cx Dilatation (cm)

Case 32 1224

Miss M.F. is a 23 year old primigravida. She is a fit lady but smokes 10-15 cigarettes per day. Her pregnancy was uneventful and she was admitted in spontaneous labour at 39 weeks on 23/3/91 at 05.40hrs. The presentation was cephalic 2/5. VE Cx 4-5cm dilated, thin. Station -2. Thin meconium-stained liquor already draining.

Labour events

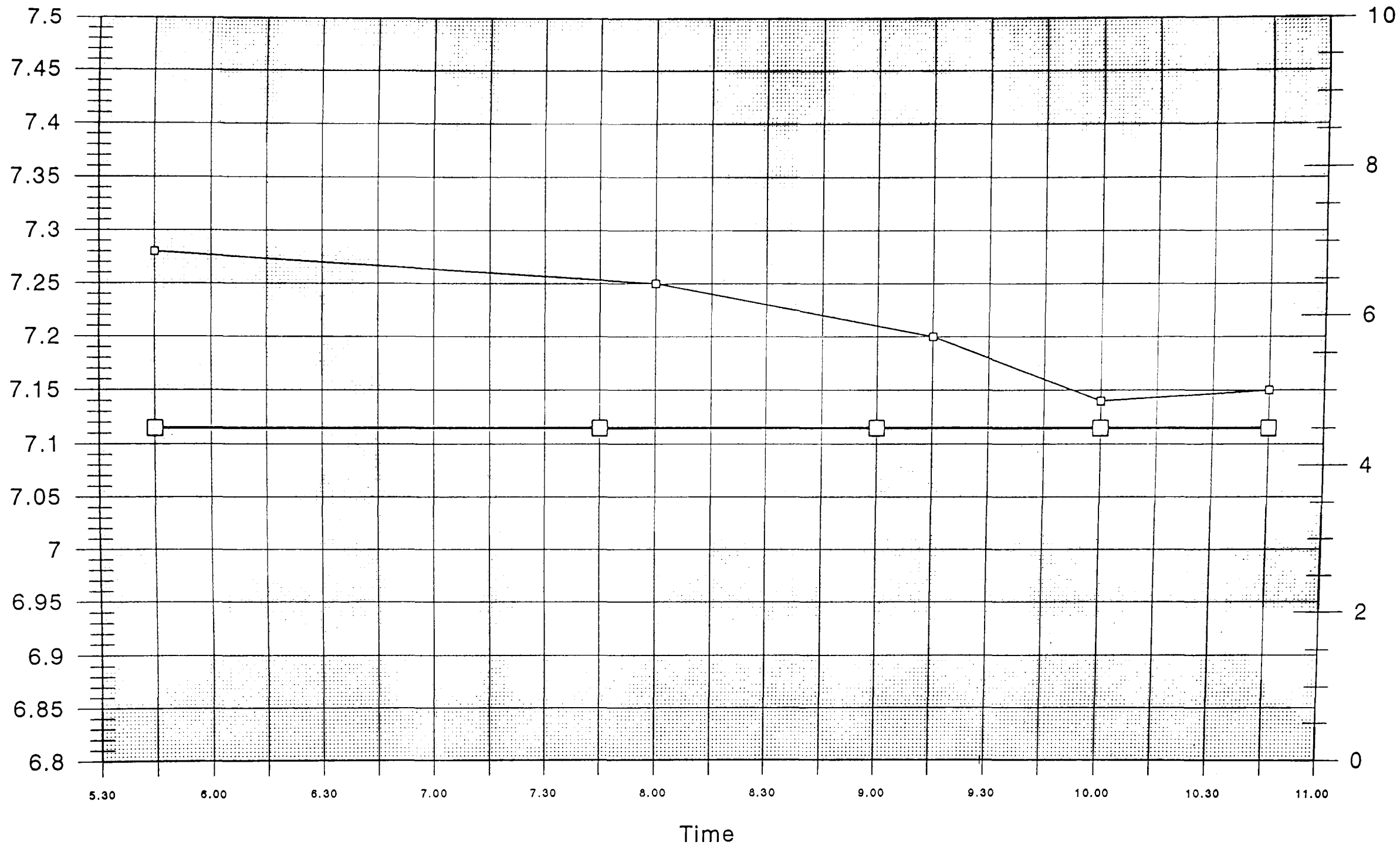
- 05.50 Pethidine 100mg, stemetil 12.5mg im.
- 06.30 Vomiting.
- 07.40 VE Cx 4-5cm dilated, thin. Station -2. FSE applied.
- 08.20 Syntocinon started.
- 09.00 VE No change.
- 10.00 VE No change.
FBS; pH 7.14 BE -14
Decision for C/S.
- 10.40 Male infant delivered by C/S under GA. Indication; fetal distress (low scalp pH at 4cm dilatation)

Outcome

Birth weight 3.33kg
Apgar 6 & 9
Cord gases pH 7.15 / 7.21 BD(ecf) 10 / 10

pH

Cx Dilatation



□ pH □ Cx Dilatation (cm)

Case 33 970

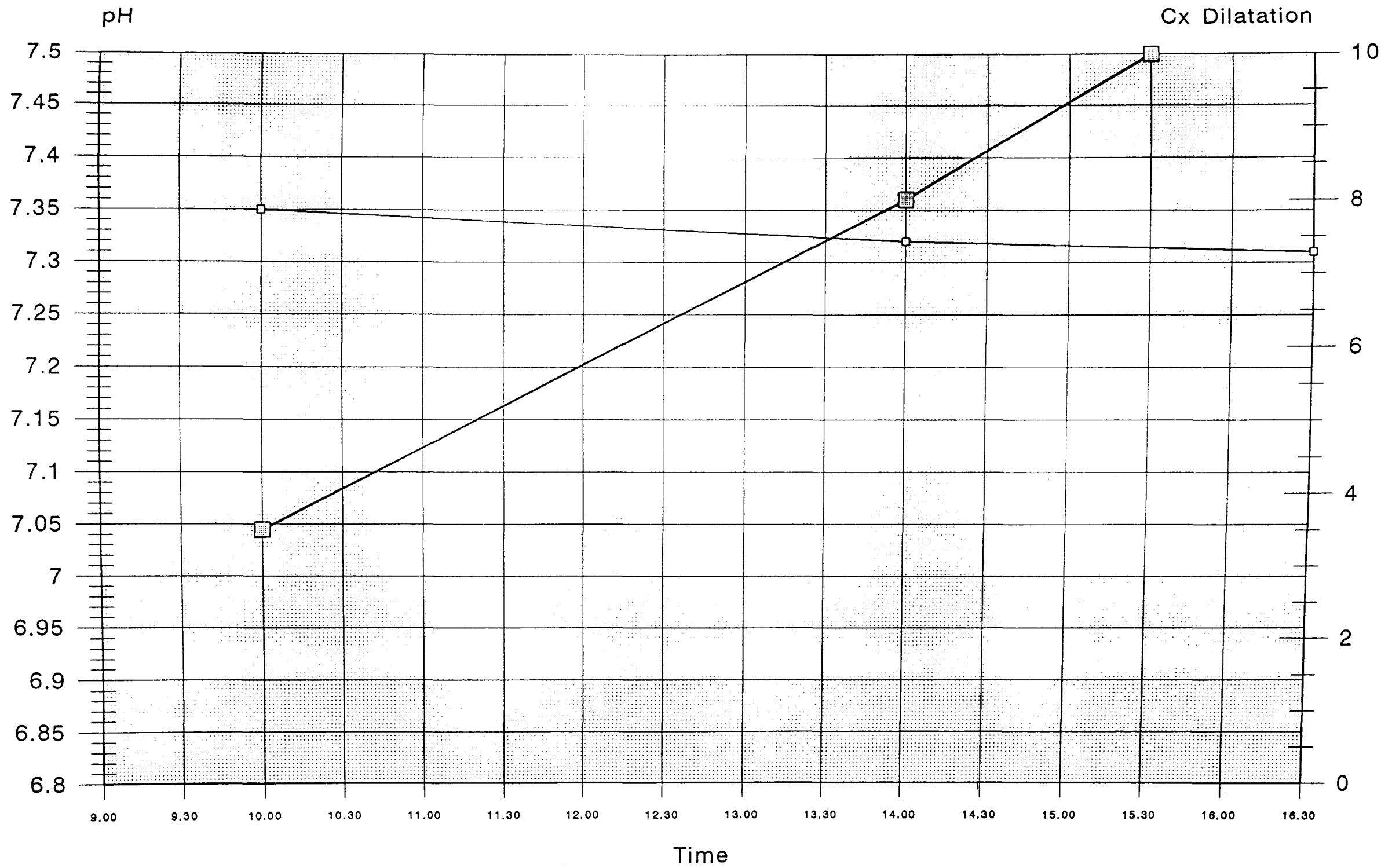
Miss S.B. is a 16 year old primigravida. She is a fit nonsmoker. There were no antenatal problems and she was admitted in spontaneous labour at 40 weeks gestation on 29.1.91. Presentation cephalic 1/5. At 09.50hrs the cervix was 3-4cm dilated; effaced and thin. Station -1. ARM was performed, and clear liquor drained.

Labour events

12.50 Pethidine 100mg, Stemetil 25mg. FSE applied.
13.45 VE Cx 7-8cm dilated. Station 0.
15.15 VE Cx fully dilated. Station +2.
15.20 Pushing began.
16.00 Episiotomy.
16.08 Normal delivery of female infant.

Outcome

Birth weight 3.63kg
Apgar 8 & 9
Cord gases pH 7.31 / 7.38 BD(ecf) 0 / 2



pH
 Cx Dilatation (cm)

Case 34 594

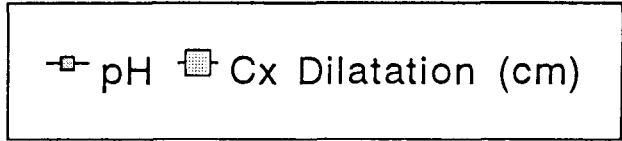
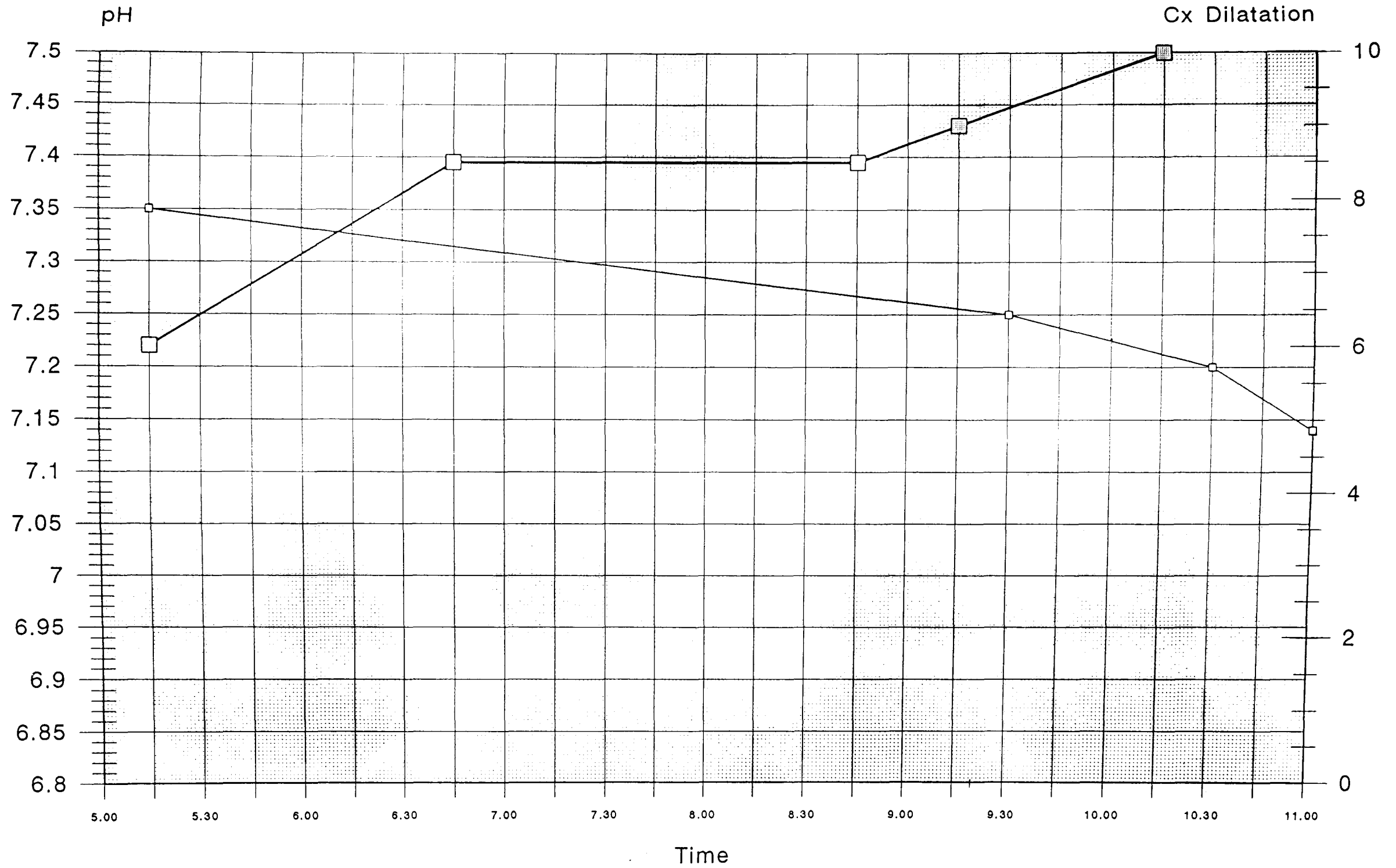
Mrs M.B. is a primigravida, 20 years of age, fit and a nonsmoker. During her pregnancy she lost 1kg in weight from booking to full term. Fetal size was felt clinically to be adequate. She was admitted in established labour at 38 weeks, at 05.15hrs on 5.11.90. Presentation cephalic 2/5; cervix already 6cm dilated, station -1. The membranes had ruptured spontaneously and clear liquor was draining; an FSE was applied.

Labour events

- 06.50 VE Cx 8-9cm dilated. Station -1.
- 08.50 VE Cx No change.
- 09.00 Pethidine 100mg im.
- 09.10 Oxygen to mother for 20 minutes.
- 09.20 VE Cx 9cm dilated. Station 0. OP position.
- 09.50 Syntocinon started.
- 10.20 VE Cx fully dilated. Station +1. OP position.
FBS; pH 7.21 BE -5.
- 10.30 Pushing begun.
- 10.55 Little progress. Meconium now present.
- 11.07 Neville Barnes forceps delivery of female infant. (Indication; slow progress, borderline scalp pH; baby delivered OP)

Outcome

Birth weight 2.19kg
Apgar 4 & 7
Cord gases pH 7.14 / 7.20 BD(ecf) 7 / 5



Case 35 1591

Mrs J.M. is a 27 year old primigravida; fit and a nonsmoker. Her pregnancy was straightforward apart from very mild hypertension at 39 weeks (BP 130/85) without proteinuria. Fetal size was judged to be adequate. Labour was induced at 41 weeks on the grounds of postmaturity and mild hypertension. She was admitted on 6.6.91 at which time the presentation was cephalic 3/5; cervix 1cm dilated, 2cm long. Prostin gel 2mg was given at 14.00hrs, and a repeat dose of 2mg was given at 18.30hrs the following day, because although the cervix was now 3cm dilated the head was still "high" and ARM felt to be inadvisable. ARM was performed at 16.35hrs on 8.6.91; cervix still 3cm dilated; station -3. Clear liquor drained, and FSE applied.

Labour events

Note; BP stable at 130/80-85 throughout labour, with no significant proteinuria.

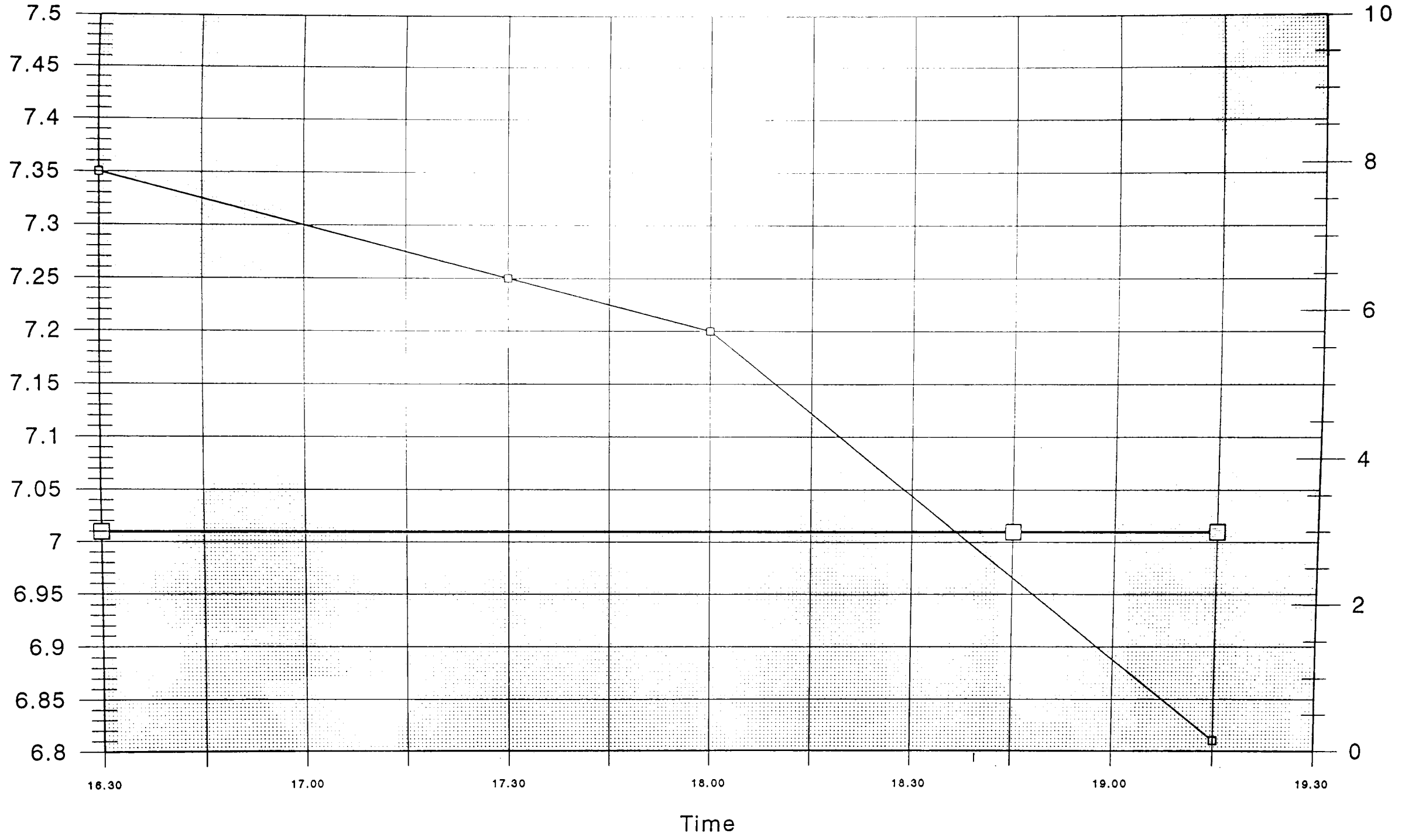
- 16.45 Maternal position changed; oxygen given for 30 minutes.
- 18.15 Epidural started.
- 18.40 VE Cx 3cm dilated. Loop of cord felt. Decision for emergency Caesarean section.
- 19.12 Male infant delivered by C/S. Indication; cord prolapse at 3cm dilatation. There was difficulty with intubation of the mother; GA was therefore abandoned and C/S was carried out under epidural.

Outcome

Birth weight 3.85kg
Apgar 3 & 8
Cord gases pH 6.81 / 7.20 BD(ecf) 11 / -

pH

Cx Dilatation



□ pH □ Cx Dilatation (cm)

Case 36 46

Mrs M.T. is a 32 year old primigravida; fit, and a nonsmoker. There were no antenatal problems. Labour was induced at 42 weeks on 10.6.90, on the grounds of postmaturity.

Labour events

- 12.15 Induction of labour. Uterus term size. Cephalic 3/5.
VE Cx 2cm dilated, partly effaced. Station -2. ARM; clear liquor. FSE applied.
- 14.15 VE difficult because of patient discomfort. Cx not reached.
- 14.55 Syntocinon started.
- 16.50 Epidural started.
- 18.40 VE Cx 5cm dilated, 75% effaced. Station -1.
- 19.05 Top up.
- 21.00 VE Cx fully dilated. Station 0. OP position. Attempt at pushing for 20 minutes.
- 21.20 No progress. Top up given to allow descent of head.
- 22.30 Maternal pyrexia 38°C (from now until delivery).
Pushing started.
- 00.00 Top up.
- 00.15 In lithotomy position.
- 00.30 Female infant delivered by Kjelland forceps. Indication; delay in 2nd stage, maternal exhaustion.

Outcome

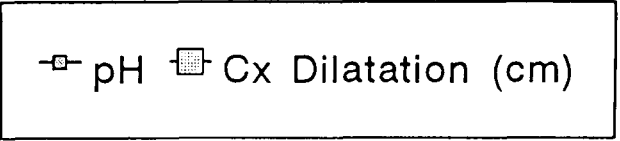
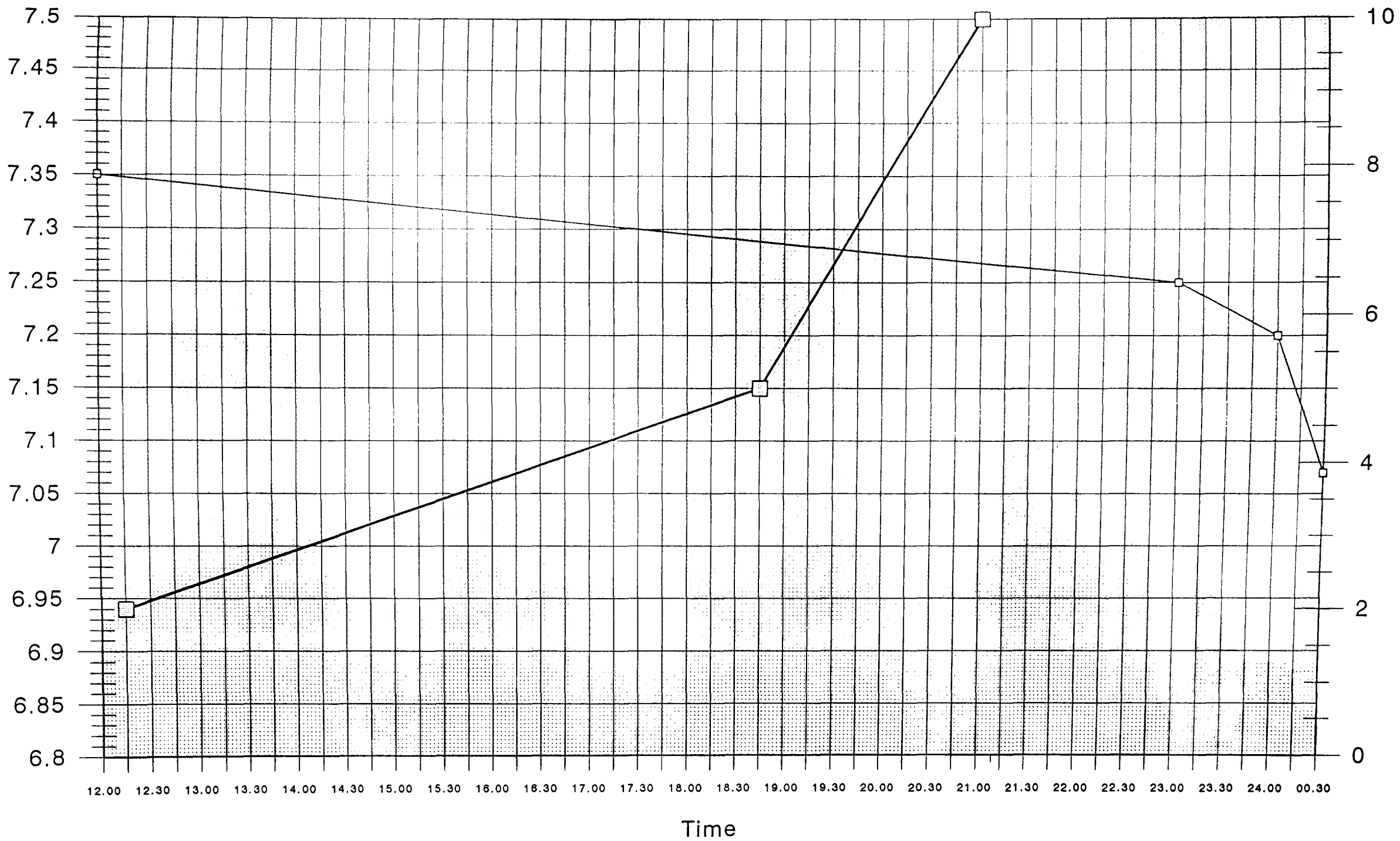
Birth weight 3.92kg

Apgar 4 & 9

Cord gases pH 7.07 / 7.30 BD(ecf) 4 / 1

pH

Cx Dilatation



Agreement results matrix for case 36

		REVIEWER																																					
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	I1	I2	J1	J2	K1	K2	L1	L2	M1	M2	N1	N2	O1	O2	P1	P2	Q1	Q2	S1	S2	R1	R2
REVIEWER	A1	100	46	66	37	88	65	74	77	69	65	84	79	51	66	87	64	64	50	48	47	71	61	89	51	48	58	57	52	72	85	71	39	44	30	47	47	28	67
	A2	46	100	38	54	52	82	52	51	69	57	55	58	86	59	56	82	81	84	66	83	61	53	56	79	85	70	70	78	61	55	61	82	36	14	73	72	25	52
	B1	66	38	100	58	74	48	74	76	50	74	76	62	38	66	71	53	49	41	47	36	62	80	71	39	37	54	43	40	62	74	63	34	87	61	49	48	22	71
	B2	37	54	58	100	40	51	36	39	49	53	44	34	46	40	43	49	51	47	38	44	37	52	43	45	45	48	39	46	43	44	37	68	62	56	43	43	12	32
	C1	88	52	74	40	100	65	84	81	67	69	88	83	54	76	90	69	63	57	55	52	82	74	91	56	53	54	59	56	70	86	84	45	57	35	52	52	26	73
	C2	65	82	48	51	65	100	51	54	81	67	67	56	88	56	70	85	96	85	64	86	59	51	71	77	86	83	68	77	75	68	61	74	35	22	73	72	24	49
	D1	74	52	74	36	84	51	100	96	54	74	80	83	54	76	75	69	51	57	68	50	81	73	76	55	53	57	59	56	56	73	83	45	58	34	66	66	31	89
	D2	77	51	76	39	81	54	96	100	58	82	77	79	52	75	77	66	56	56	64	50	76	70	77	53	51	67	60	55	62	72	77	44	58	35	65	65	30	90
	E1	69	69	50	49	67	81	54	58	100	83	70	61	73	61	74	68	83	86	73	66	62	53	75	88	71	77	80	88	79	72	62	73	37	25	83	83	26	52
	E2	65	57	74	53	69	67	74	82	83	100	74	60	59	69	74	62	70	68	75	55	63	62	74	69	57	77	72	71	73	73	64	58	55	43	80	80	25	71
	F1	84	55	76	44	88	67	80	77	70	74	100	80	59	83	88	68	65	59	59	54	81	76	89	61	57	57	63	60	82	93	81	47	61	36	55	55	26	69
	F2	79	58	62	34	83	56	83	79	61	60	80	100	65	85	85	72	55	63	61	59	95	81	86	66	62	46	71	65	62	79	94	48	56	22	58	58	32	82
	G1	51	86	38	46	54	88	54	52	73	59	59	65	100	65	59	88	89	91	68	88	66	57	60	88	96	76	76	86	63	60	64	78	36	14	79	79	27	52
	G2	66	59	66	40	76	56	76	75	61	69	83	85	65	100	81	69	56	63	68	57	89	84	81	66	61	50	76	66	78	88	86	49	67	31	60	60	28	73
	H1	87	56	71	43	90	70	75	77	74	74	88	85	59	81	100	66	70	61	61	58	88	79	99	62	57	60	66	63	78	89	87	47	57	32	56	56	29	75
	H2	64	82	53	49	69	85	69	66	68	62	68	72	88	69	66	100	84	85	64	86	70	61	66	77	85	74	69	78	62	67	70	74	47	21	73	72	25	65
	I1	64	81	49	51	63	96	51	56	83	70	65	55	89	56	70	84	100	87	64	83	57	50	70	78	87	88	69	79	76	67	57	75	37	24	75	75	24	50
	I2	50	84	41	47	57	85	57	56	86	68	59	63	91	63	61	85	87	100	72	86	65	57	61	95	89	80	77	96	68	59	67	84	39	15	88	88	27	56
	J1	48	66	47	38	55	64	68	64	73	75	59	61	68	68	61	64	64	72	100	65	65	56	61	72	66	60	90	71	77	63	65	56	41	30	80	79	32	66
	J2	47	83	36	44	52	86	50	50	66	55	54	59	88	57	58	86	83	86	65	100	63	54	58	79	91	73	69	80	62	54	64	74	35	13	72	73	26	53
	K1	71	61	62	37	82	59	81	76	62	63	81	95	66	89	88	70	57	65	65	63	100	88	89	70	64	46	71	69	64	80	97	50	59	23	59	59	32	81
	K2	61	53	80	52	74	51	73	70	53	62	76	81	57	84	79	61	50	57	56	54	88	100	79	60	54	41	61	59	61	73	86	44	87	40	51	51	27	71
	L1	89	56	71	43	91	71	76	77	75	74	89	86	60	81	99	66	70	61	61	58	89	79	100	63	58	59	67	63	77	90	88	48	57	32	56	56	29	73
	L2	51	79	39	45	56	77	55	53	88	69	61	66	88	66	62	77	78	95	72	79	70	60	63	100	84	71	78	98	66	61	67	80	36	14	88	88	28	56
	M1	48	85	37	45	53	86	53	51	71	57	57	62	96	61	57	85	87	89	66	91	64	54	58	84	100	74	72	81	60	56	65	76	34	13	78	79	26	52
	M2	58	70	54	48	54	83	57	67	77	77	57	46	76	50	60	74	88	80	60	73	46	41	59	71	74	100	60	75	68	58	47	70	35	29	79	79	23	60
	N1	57	70	43	39	59	68	59	60	80	72	63	71	76	76	66	69	69	77	90	69	71	61	67	78	72	60	100	78	83	72	71	59	44	24	72	71	29	59
	N2	52	78	40	46	56	77	56	55	88	71	60	65	86	66	63	78	79	96	71	80	69	59	63	98	81	75	78	100	68	61	67	80	37	15	87	87	28	58
	O1	72	61	62	43	70	75	56	62	79	73	82	62	63	78	78	62	76	68	77	62	64	61	77	66	60	68	83	68	100	87	64	53	53	35	62	62	27	58
	O2	85	55	74	44	86	68	73	72	72	73	93	79	60	88	89	67	67	59	63	54	80	73	90	61	56	58	72	61	87	100	79	47	62	39	56	56	26	66
	P1	71	61	63	37	84	61	83	77	62	64	81	94	64	86	87	70	57	67	65	64	97	86	88	67	65	47	71	67	64	79	100	50	60	23	60	60	31	81
	P2	39	82	34	68	45	74	45	44	73	58	47	48	78	49	47	74	75	84	56	74	50	44	48	80	76	70	59	80	53	47	50	100	32	13	76	75	21	42
	Q1	44	36	87	62	57	35	58	58	37	55	61	56	36	67	57	47	37	39	41	35	59	87	57	36	34	35	44	37	53	62	60	32	100	60	36	36	18	53
	Q2	30	14	61	56	35	22	34	35	25	43	36	22	14	31	32	21	24	15	30	13	23	40	32	14	13	29	24	15	35	39	23	13	60	100	22	22	22	30
S1	47	73	49	43	52	73	66	65	83	80	55	58	79	60	56	73	75	88	80	72	59	51	56	88	78	79	72	87	62	56	60	76	36	22	100	100	29	66	
S2	47	72	48	43	52	72	66	65	83	80	55	58	79	60	56	72	75	88	79	73	59	51	56	88	79	79	71	87	62	56	60	75	36	22	100	100	29	66	
R1	28	25	22	12	26	24	31	30	26	25	26	32	27	28	29	25	24	27	32	26	32	27	29	28	26	23	29	28	27	26	31	21	18	22	29	29	100	33	
R2	67	52	71	32	73	49	89	90	52	71	69	82	52	73	75	65	50	56	66	53	81	71	73	56	52	60	59	58	58	66	81	42	53	30	66	66	33	100	

Case 37 894

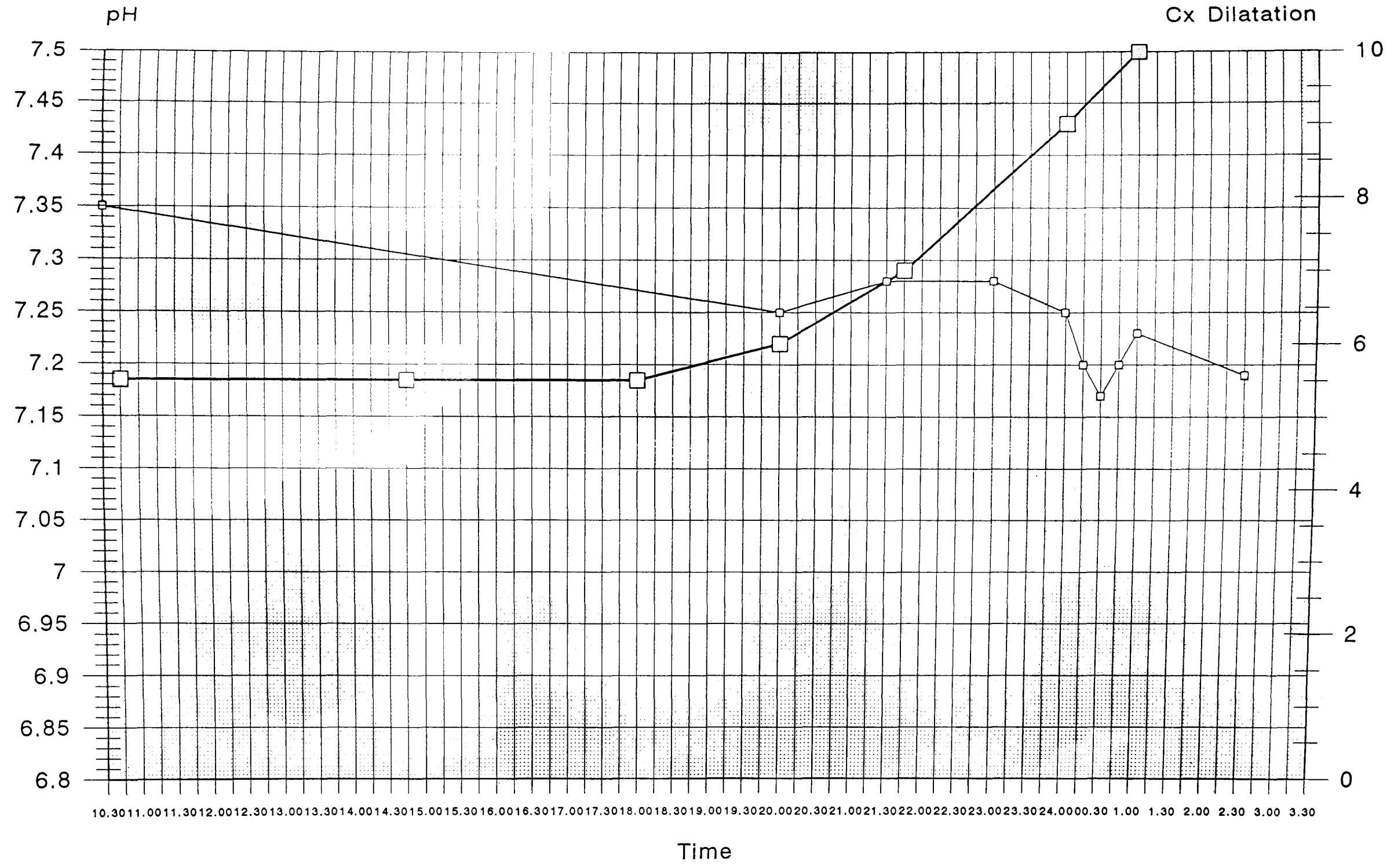
Miss M.L. is a 17 year old primigravida. She smokes 5 cigarettes per day and has mild asthma, for which she uses no regular medication. Her pregnancy was straightforward and she laboured spontaneously at 40 weeks. When she was admitted on 30.12.90 at 10.50hrs the fetal head was 2/5 palpable, and the cervix 5-6cm dilated. ARM was performed and clear liquor drained; an FSE was applied.

Labour events

11.04 Bedpan.
14.50 VE No change.
15.30 Syntocinon begun.
15.40 Epidural begun.
16.00 Epidural complete.
17.50 Bedpan.
18.00 VE Cx No change.
18.15 Top up.
20.05 VE Cx 6cm dilated. OP position. IUPD inserted.
FBS; pH 7.25 BE -1
20.30 Top up.
21.50 VE Cx 7cm dilated. Station -1. OP position. Caput ++ moulding ++.
22.00 Top up.
00.05 VE Cx 9cm dilated. Station 0. OP position. Caput ++ moulding ++.
Top up.
00.07 Syntocinon stopped. Maternal oxygen.
00.25 Syntocinon restarted.
01.00 VE Cx fully dilated. Station +2. Uncertain position.
01.25 Pushing begun.
02.05 Top up.
02.30 Delivery of male infant by Neville Barnes forceps. Indication; maternal exhaustion.

Outcome

Birth weight 2.95kg
Apgar 8 & 9
Cord gases pH 7.19 / 7.22 BD(ecf) 7 / 6



○ pH
 □ Cx Dilatation (cm)

Agreement results matrix for case 37

		REVIEWER																																					
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	I1	I2	J1	J2	K1	K2	L1	L2	M1	M2	N1	N2	O1	O2	P1	P2	Q1	Q2	S1	S2	R1	R2
REVIEWER	A1	100	43	90	47	48	50	93	53	93	96	47	34	43	56	53	51	41	47	55	55	41	38	51	51	44	46	47	48	49	56	50	44	59	77	51	46	36	37
	A2	43	100	42	66	81	68	48	68	40	44	72	69	75	84	38	74	77	83	64	66	73	71	66	75	90	80	75	64	71	79	68	76	72	30	80	84	70	65
	B1	90	42	100	56	51	55	91	53	89	94	48	36	44	48	60	48	43	51	51	48	45	40	48	48	46	45	47	47	53	53	49	44	51	79	47	44	36	37
	B2	47	66	56	100	78	92	48	87	49	48	82	75	80	63	47	85	77	68	84	85	90	81	88	87	61	65	80	79	92	65	80	81	58	49	57	58	46	45
	C1	48	81	51	78	100	79	49	76	51	51	69	62	67	75	43	73	72	81	69	73	74	68	75	73	71	80	72	66	81	83	70	67	72	38	77	80	60	57
	C2	50	68	55	92	79	100	52	90	52	54	91	83	87	62	46	85	78	70	86	84	90	90	92	86	62	65	84	83	95	66	88	87	57	39	62	64	50	51
	D1	93	48	91	48	49	52	100	50	90	94	51	41	50	52	46	52	45	46	50	50	47	45	52	52	52	51	46	50	49	56	47	49	54	71	52	49	41	43
	D2	53	68	53	87	76	90	50	100	53	56	82	76	82	64	46	85	77	74	90	83	86	78	91	88	59	60	84	76	94	67	85	85	57	40	56	58	54	47
	E1	93	40	89	49	51	52	90	53	100	93	47	37	43	48	71	50	40	50	57	51	43	37	51	50	44	46	48	45	52	56	52	43	49	65	46	42	36	35
	E2	96	44	94	48	51	54	94	56	93	100	47	38	45	51	55	51	44	53	57	51	44	40	52	51	46	47	48	48	53	57	52	46	53	71	49	46	41	38
	F1	47	72	48	82	69	91	51	82	47	47	100	91	92	62	38	89	82	66	84	79	90	91	87	92	67	60	86	78	85	65	88	91	56	36	67	69	53	54
	F2	34	69	36	75	62	83	41	76	37	38	91	100	93	61	37	80	75	58	82	76	83	90	84	83	65	60	83	74	79	60	90	91	49	20	65	70	58	51
	G1	43	75	44	80	67	87	50	82	43	45	92	93	100	68	36	86	79	64	85	78	88	92	89	90	73	68	90	83	85	65	93	98	61	28	68	71	55	55
	G2	56	84	48	63	75	62	52	64	48	51	62	61	68	100	41	68	70	75	66	74	60	62	65	66	74	70	67	67	64	68	65	70	93	50	87	84	57	55
	H1	53	38	60	47	43	46	46	46	71	55	38	37	36	41	100	39	37	49	51	43	38	33	40	39	40	38	41	33	47	48	47	37	36	45	31	33	38	32
	H2	51	74	48	85	73	85	52	85	50	51	89	80	86	68	39	100	91	69	87	87	89	82	88	96	71	67	91	80	86	72	86	87	62	38	63	65	50	52
	I1	41	77	43	77	72	78	45	77	40	44	82	75	79	70	37	91	100	74	77	83	83	76	76	83	69	61	85	70	79	61	72	80	63	34	67	70	50	49
	I2	47	83	51	68	81	70	46	74	50	53	66	58	64	75	49	69	74	100	70	61	68	63	63	70	74	73	68	61	72	77	65	65	68	38	72	76	67	62
	J1	55	64	51	84	69	86	50	90	57	57	84	82	85	66	51	87	77	70	100	82	81	76	92	89	58	59	86	74	87	68	92	86	55	39	56	58	52	46
	J2	55	66	48	85	73	84	50	83	51	51	79	76	78	74	43	87	83	61	82	100	85	79	88	84	61	66	81	76	85	65	77	80	68	49	64	63	49	45
	K1	41	73	45	90	74	90	47	86	43	44	90	83	88	60	38	89	83	68	81	85	100	90	88	92	69	67	84	84	92	63	80	86	55	33	63	66	51	49
	K2	38	71	40	81	68	90	45	78	37	40	91	90	92	62	33	82	76	63	76	79	90	100	85	84	67	68	84	87	85	58	86	91	56	25	68	70	55	54
	L1	51	66	48	88	75	92	52	91	51	52	87	84	89	65	40	88	76	63	92	88	88	85	100	90	62	67	85	82	89	69	90	89	56	37	59	61	51	48
	L2	51	75	48	87	73	86	52	88	50	51	92	83	90	66	39	96	83	70	89	84	92	84	90	100	70	66	88	81	88	73	88	89	61	38	64	66	52	51
	M1	44	90	46	61	71	62	52	59	44	46	67	65	73	74	40	71	69	74	58	61	69	67	62	70	100	88	74	62	65	81	64	72	63	29	71	74	71	68
	M2	46	80	45	65	80	65	51	60	46	47	60	60	68	70	38	67	61	73	59	66	67	68	67	66	88	100	69	66	68	87	64	66	64	29	71	74	70	68
	N1	47	75	47	80	72	84	46	84	48	48	86	83	90	67	41	91	85	68	86	81	84	84	85	88	74	69	100	77	89	71	92	90	61	30	65	67	51	51
	N2	48	64	47	79	66	83	50	76	45	48	78	74	83	67	33	80	70	61	74	76	84	87	82	81	62	66	77	100	82	54	78	82	70	38	68	67	46	48
	O1	49	71	53	92	81	95	49	94	52	53	85	79	85	64	47	86	79	72	87	85	92	85	89	88	65	68	89	82	100	69	87	85	58	38	60	62	49	49
	O2	56	79	53	65	83	66	56	67	56	57	65	60	65	68	48	72	61	77	68	65	63	58	69	73	81	87	71	54	69	100	70	65	62	38	70	73	74	68
	P1	50	68	49	80	70	88	47	85	52	52	88	90	93	65	47	86	72	65	92	77	80	86	90	88	64	64	92	78	87	70	100	92	56	31	64	66	53	49
	P2	44	76	44	81	67	87	49	85	43	46	91	91	98	70	37	87	80	65	86	80	86	91	89	89	72	66	90	82	85	65	92	100	63	30	67	69	57	55
	Q1	59	72	51	58	72	57	54	57	49	53	56	49	61	93	36	62	63	68	55	68	55	56	56	61	63	64	61	70	58	62	56	63	100	55	91	88	49	54
	Q2	77	30	79	49	38	39	71	40	65	71	36	20	28	50	45	38	34	38	39	49	33	25	37	38	29	29	30	38	38	31	30	55	100	43	37	23	27	
S1	51	80	47	57	77	62	52	56	46	49	67	65	68	87	31	63	67	72	56	64	63	68	59	64	71	71	65	68	60	70	64	67	91	43	100	99	58	62	
S2	46	84	44	58	80	64	49	58	42	46	69	70	71	84	33	65	70	76	58	63	66	70	61	66	74	74	67	67	62	73	66	69	88	37	99	100	63	64	
R1	36	70	36	46	60	50	41	54	36	41	53	58	55	57	38	50	50	67	52	49	51	55	51	52	71	70	51	46	49	74	53	57	49	23	58	63	100	68	
R2	37	65	37	45	57	51	43	47	35	38	54	51	55	55	32	52	49	62	46	45	49	54	48	51	68	68	51	48	49	68	49	55	54	27	62	64	68	100	

Case 38 2009

Mrs D.P. is a 24 year old primigravida. She is a fit lady but smokes 6 cigarettes per day. There were no antenatal problems, and spontaneous labour occurred at 37 weeks. She was admitted at 14.35hrs on 9.9.91, and at this time the presentation was cephalic 2/5, cervix 2-3cm dilated, effacing.

Labour events

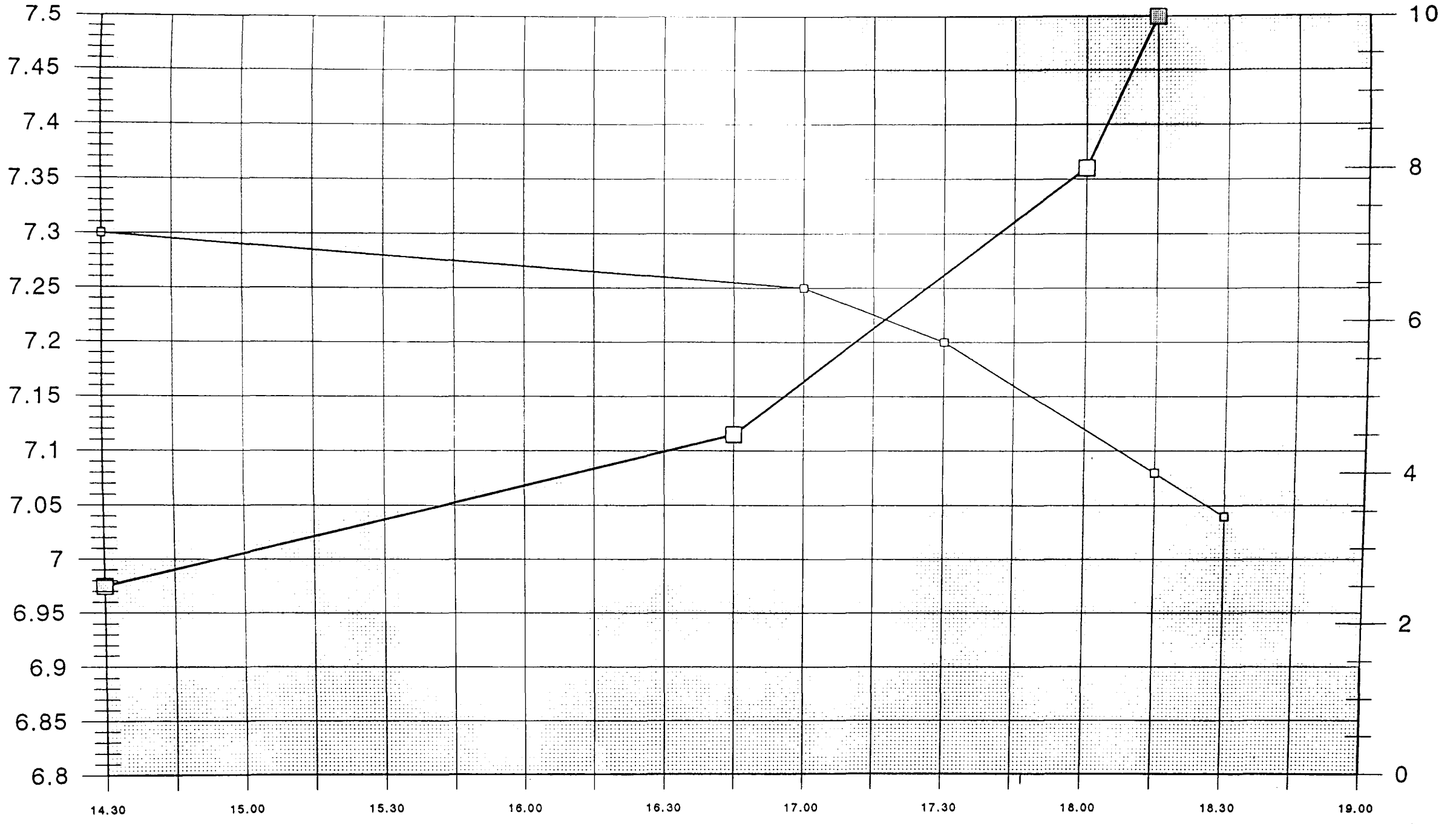
- 15.00 Enema given.
- 15.30 Bath.
- 16.40 VE Cx 4-5cm dilated, thinning. Station -1. ARM; liquor stained with fresh meconium. FSE applied.
- 16.50 Pethidine 100mg, phenergan 25mg im.
- 17.55 VE Cx 8cm dilated. Station +1. OA position.
- 18.05 Maternal oxygen given.
- 18.15 FBS; pH 7.08 BE -11
- 18.20 VE Cx fully dilated. Pushing begun.
- 18.30 Episiotomy.
- 18.33 Normal delivery of male infant.

Outcome

Birth weight 2.61kg
Apgar 3 & 6 (9 at 10 minutes)
Cord gases pH 7.04 / 7.08 BD(ecf) 11 / 11

pH

Cx Dilatation



Time



Case 39 1305

Mrs K.G. is a 22 year old primigravida; fit and a nonsmoker. Her pregnancy was entirely straightforward and she was admitted in spontaneous labour at 40 weeks gestation on 3.4.91. The uterus was term size, presentation cephalic 2/5, and on VE at 02.50hrs the cervix was found to be closed. Contractions continued and the membranes ruptured spontaneously at 04.15 with the drainage of clear liquor. By 07.45hrs the cervix was 3-4cm dilated, 1cm thick. Station -1. OP position.

Labour events

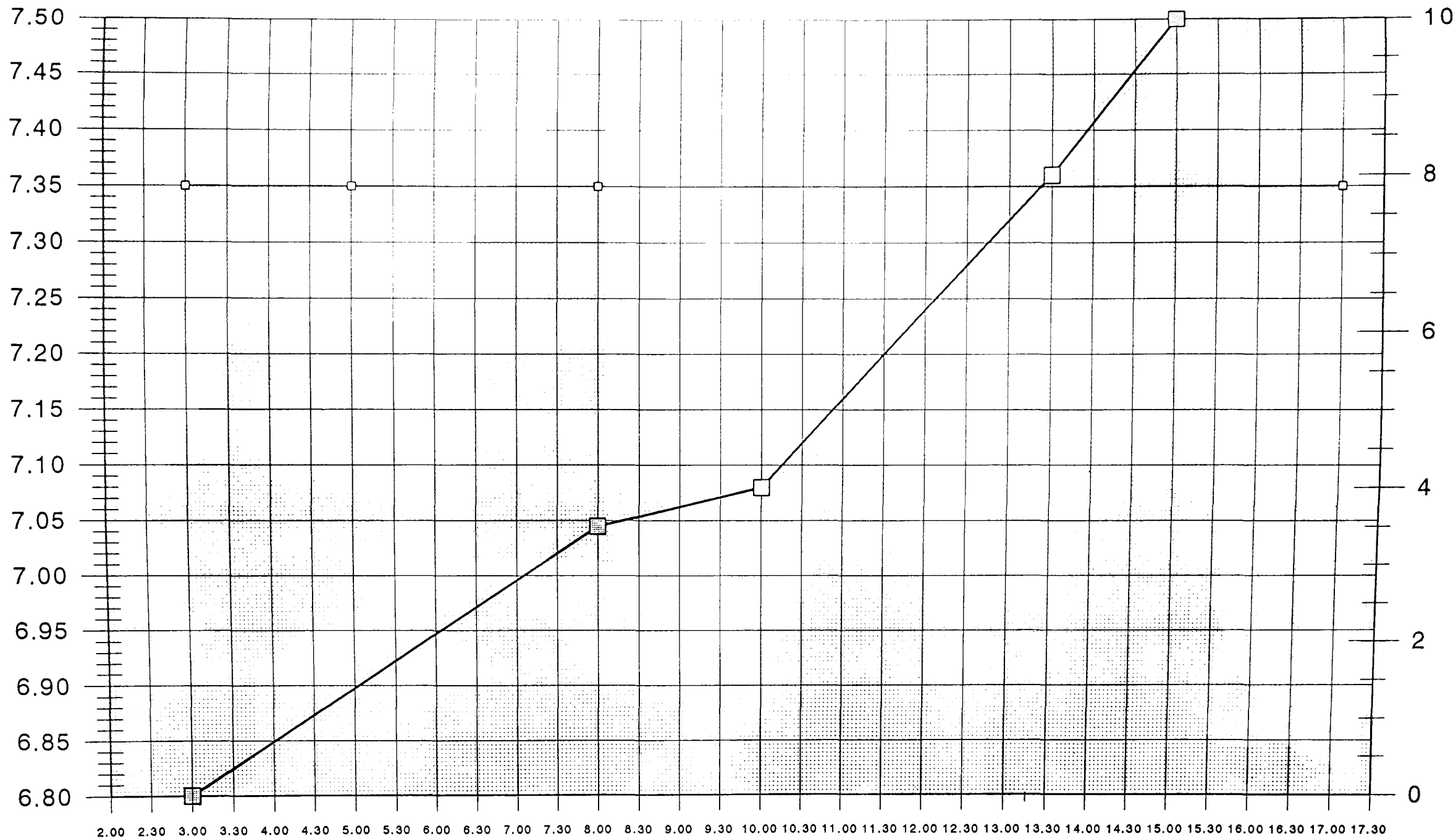
10.00 VE Cx 4cm dilated, 1cm thick.
11.40 Epidural inserted.
11.55 Syntocinon started.
13.30 VE Cx 8cm dilated. Station -1. OP position.
13.45 Top up.
15.15 VE Cx fully dilated. Station 0. LOP position.
15.55 Pushing began.
17.00 Episiotomy.
17.05 Normal delivery of female infant.

Outcome

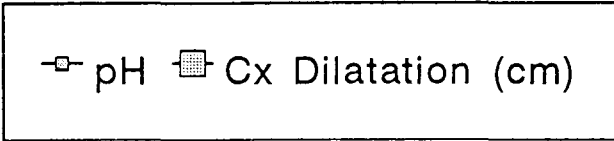
Birth weight 3.59kg
Apgar 9 & 9
Cord gases pH 7.38 / 7.40 BD(ecf) 8 / -6

pH

Cx Dilatation



Time



Case 40 1688

Miss K.W. is a 17 year old primigravida. She is generally well, but smokes 5 cigarettes per day. No problems were detected antenatally, and intrauterine growth retardation was not suspected until she was admitted at 37 weeks with reduced fetal movements. At that time it was noted that the fundal height measured only 31cm. Moreover, the CTG was abnormal and labour was induced straight away. (BP was normal and there was no proteinuria.)

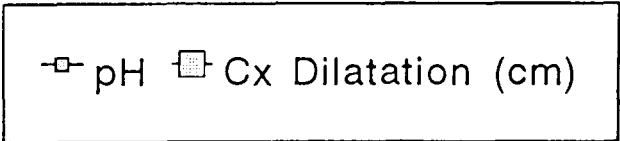
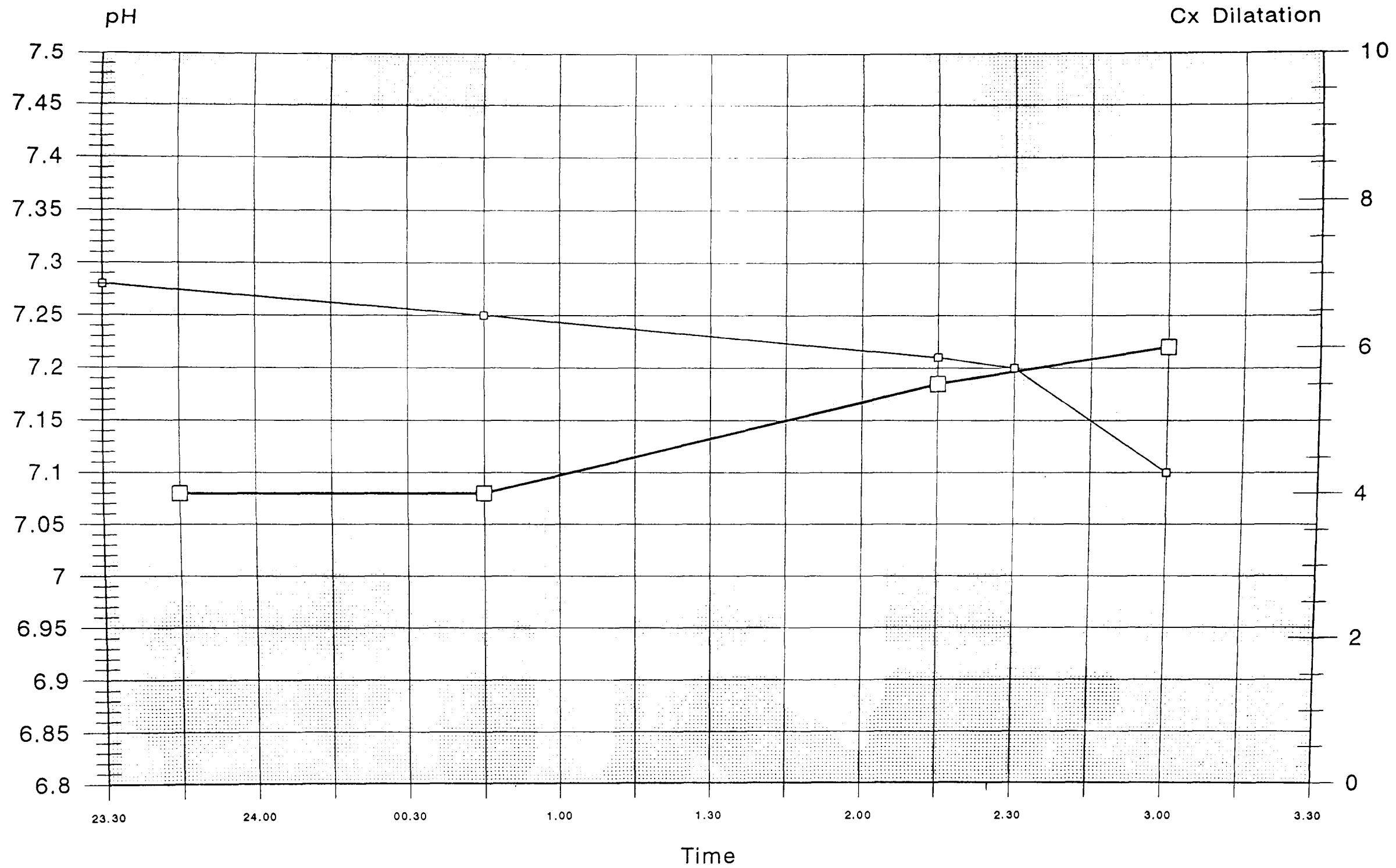
Labour events

3.7.91

- 00.40 Induction of labour. Cephalic 3/5. VE Cx 4cm dilated, 1cm thick. Station -1. ARM; clear liquor. FSE applied
- 02.00 Bed pan..
- 02.20 VE Cx 5-6cm dilated. FBS; pH 7.21 BE -3
Decision for Caesarean section.
- 03.06 Male infant delivered by C/S under GA. Indication; low fetal scalp pH early in labour; baby thought to be growth retarded.

Outcome

Birth weight 1.96kg
Apgar 7 & 9
Cord gases pH 7.25 BD(ecf) 3 / 6



Case 41 2232

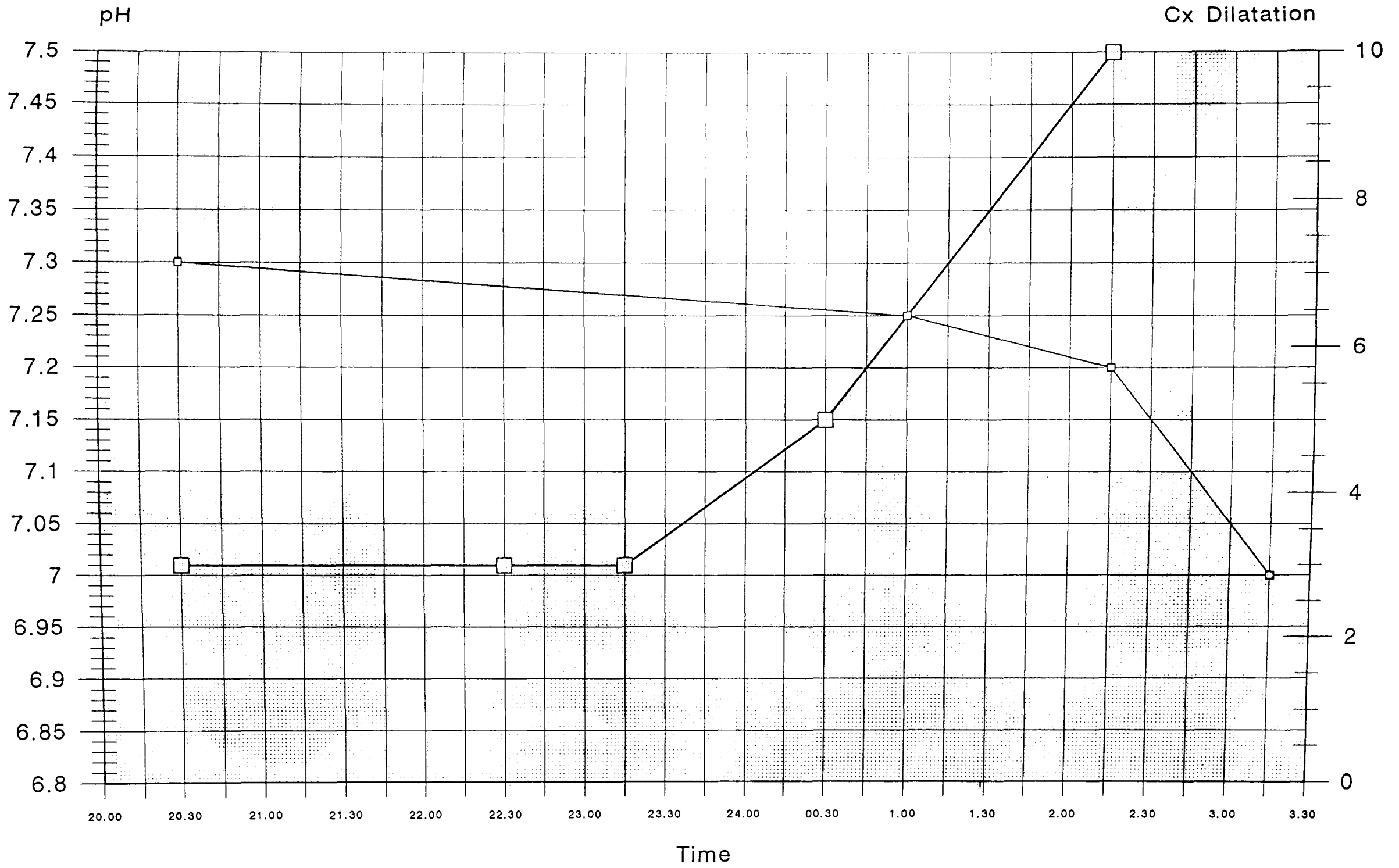
Miss J.S. is a 19 year old primigravida. She has asthma for which she uses a ventolin inhaler on a regular basis. She is a nonsmoker. Her pregnancy was entirely uneventful and spontaneous labour occurred at 41 weeks. She was admitted on 24.10.91 at 07.30hrs in very early labour, and the cervix was then just 1cm dilated, partly effaced (head 2/5 palpable). She returned to the labour ward at 20.00hrs, when labour was becoming established.

Labour events

- 20.30 VE Cx 3cm dilated, thin. Station -1.
- 20.45 Pethidine 100mg, phenergan 25mg im.
- 22.30 VE Cx 3cm dilated, thin. Station -1. ARM; clear liquor.
- 23.10 VE Cx No change. FSE applied.
- 23.50 Epidural begun
- 00.30 VE Cx 5cm dilated. Station -1. OP position.
- 01.00 Contractions felt to be mild and only 2:10. Syntocinon started.
- 01.50 Top up.
- 02.10 VE Cx fully dilated. Station +1. OA position.
- 02.30 Pushing begun.
- 03.11 Normal delivery of female infant.

Outcome

Birth weight 2.84kg
Apgar 9 & 9
Cord gases pH 6.97 / 7.06 BD(ecf) 12 / 7



□ pH □ Cx Dilatation (cm)

Case 42 1303

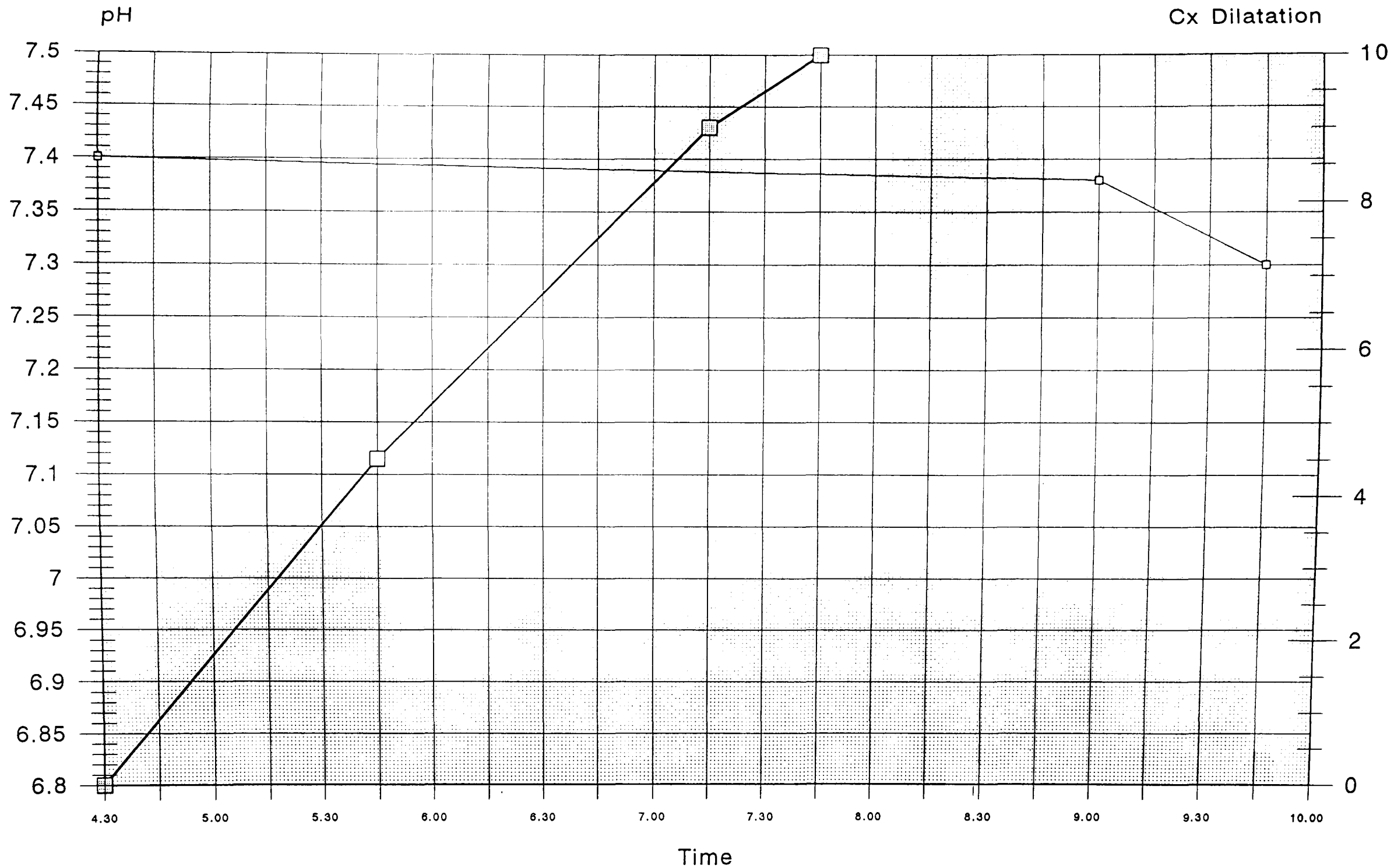
Mrs S.E. is a 32 year old primigravida. She is a fit lady and does not smoke. There were no antenatal problems and she was admitted in spontaneous labour at 41 weeks on 3.4.91 at 00.20hrs. The membranes had recently ruptured, but she had had no contractions. The uterus was term size, cephalic 1/5. VE (01.15hrs) Cx closed, uneffaced. Station -2. Liquor stained with meconium.

Labour events

- 04.30 VE No change. Stemetil 12.5mg im.
- 05.17 Epidural inserted.
- 05.40 VE Cx 4-5cm dilated, effaced. Station -2. FSE applied.
- 07.10 VE Cx 9cm dilated. Station -2. ROP position.
- 07.30 Top up.
- 07.52 VE Cx fully dilated. Station 0. ROP position.
- 08.10 Pushing begun.
- 08.40 Syntocinon started.
- 09.00 VE Cx fully dilated. Station +1. OP position. Caput + moulding +.
FBS pH 7.38
- 09.10 Top up.
- 09.20 Decision for forceps delivery
- 09.30 In dorsal position for catheterisation.
- 09.46 Female infant delivered by Neville Barnes forceps. Indication; prolonged fetal bradycardia in 2nd stage. (Delivered in OA position)

Outcome

Birth weight 3.05kg
Apgar 9 & 9
Cord gases pH 7.29 / 7.34 BD(ecf) -2 / -1



□ pH □ Cx Dilatation (cm)

Case 43 631

Miss T.W. is a 22 year old lady expecting her second baby. Her first child was delivered in 1987 by emergency Caesarean section because of fetal distress prior to established labour. The child was born at term and weighed 7lb 6oz. During this pregnancy no problems were detected antenatally, and it was planned that she should have a trial of vaginal delivery. She laboured spontaneously at 40 weeks gestation, being admitted at 08.30hrs on 13.11.90. Presentation cephalic 3/5. Cervix 2cm dilated, thick; station -1. Continuous monitoring was begun at 10.20hrs.

Labour events

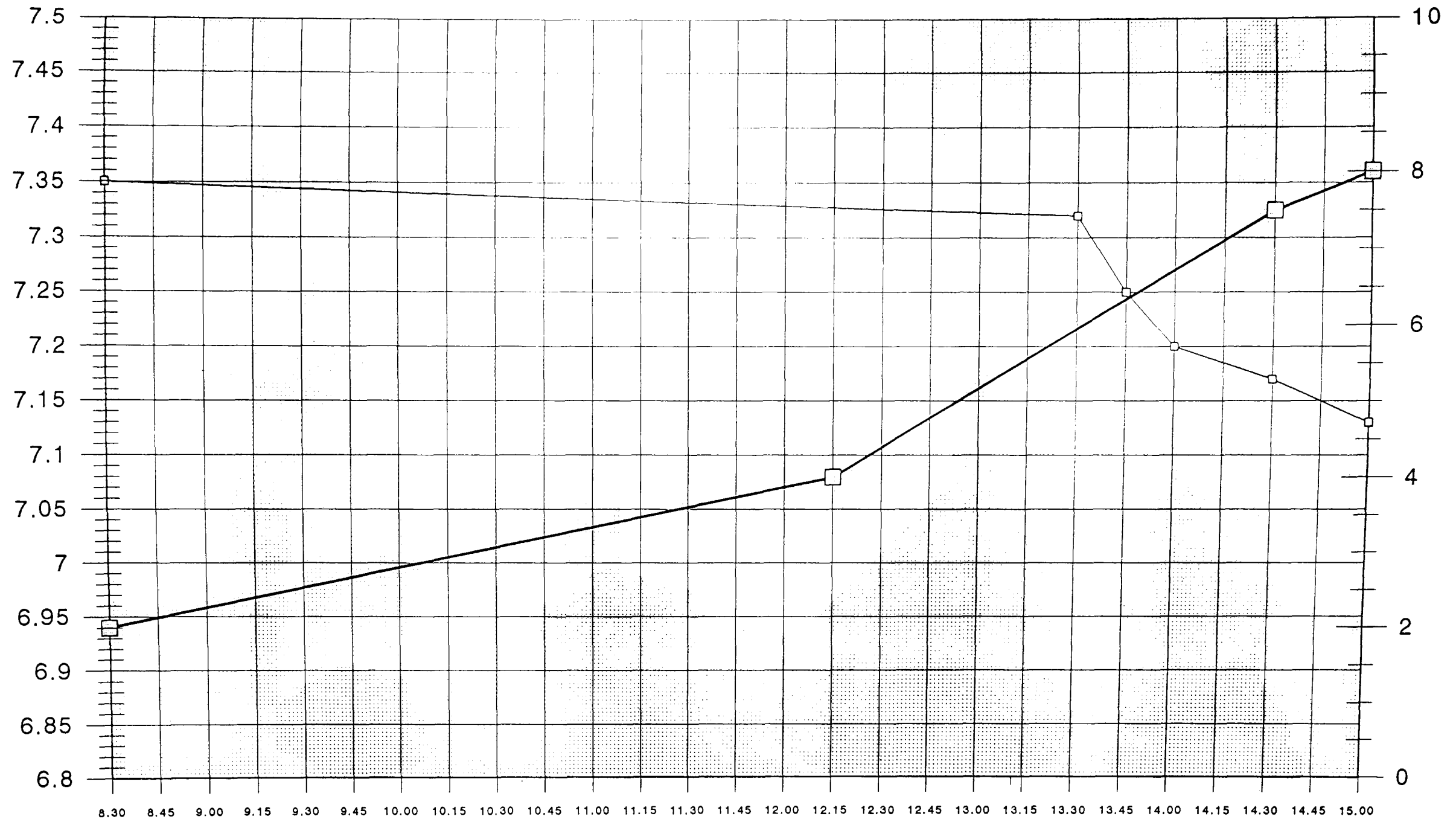
- 11.00 Pethidine 50mg, Stemetil 12.5mg im.
- 11.10 Vomited.
- 12.15 VE Cx 4cm dilated, thin. Station -2. ARM; clear liquor. FSE applied.
- 13.10 Epidural begun.
- 13.45 Epidural main dose.
- 14.25 VE Cx 7-8cm dilated. Station -1.
FBS; pH 7.17 BE-7
Decision for Caesarean section.
- 15.00 Male infant delivered by emergency C/S under GA. Indication; fetal distress.
Cord found alongside fetal head at operation.

Outcome

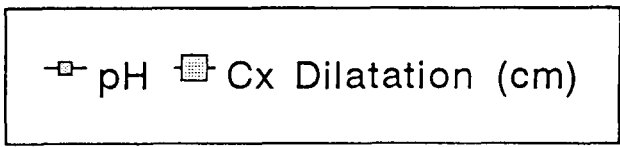
Birth weight 3.66kg
Apgar 7 & 9
Cords pH 7.13 / 7.17 BD(ecf) 5 / 5

pH

Cx Dilatation



Time



Case 44 2129

Mrs L.C. is a 34 year old lady expecting her third baby. She is fit and well, but smokes 10 cigarettes per day. Her two previous children were born normally at term, and weighed 8lb 7oz and 9lb respectively. Her current pregnancy was straightforward until the development of mild hypertension at 38 weeks. By 39 weeks her BP was 150/100 and there was 2+ proteinuria. She was asymptomatic and reflexes were normal. The fetus was felt to be slightly small for dates (fundal height 37cm). Labour was induced at this stage on the grounds of proteinuric hypertension. Presentation cephalic 2/5. VE Cx 1cm dilated, 1cm long. Prostin tablet 3mg was given PV at 06.45hrs on 3.10.91. ARM at 12.00hrs; clear liquor. FSE applied.

Labour events

Note; maternal condition gave rise to no concern during labour.

- 12.30 Syntocinon begun.
- 15.30 VE Cx 2cm dilated, thick. Station -2.
- 16.40 Decision for C/S
- 17.30 Male infant delivered by emergency C/S under GA. Indication; fetal tachycardia with decelerations at 2cm dilatation.

Outcome

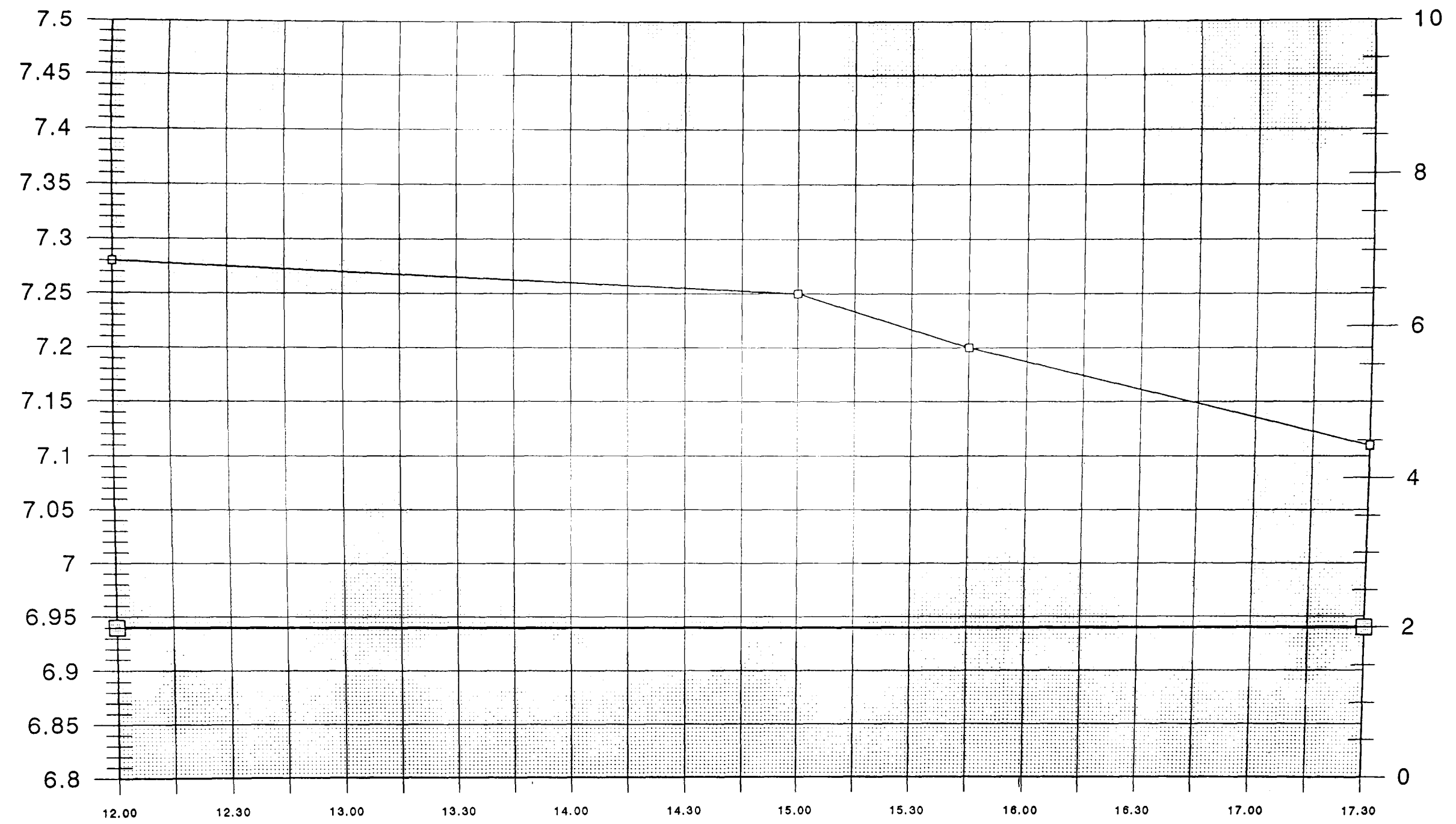
Birth weight 2.13kg

Apgar 6 & 9

Cord gases pH 7.11 / 7.15 BD(ecf) 8 / 7

pH

Cx Dilatation



Time



Case 45 2337

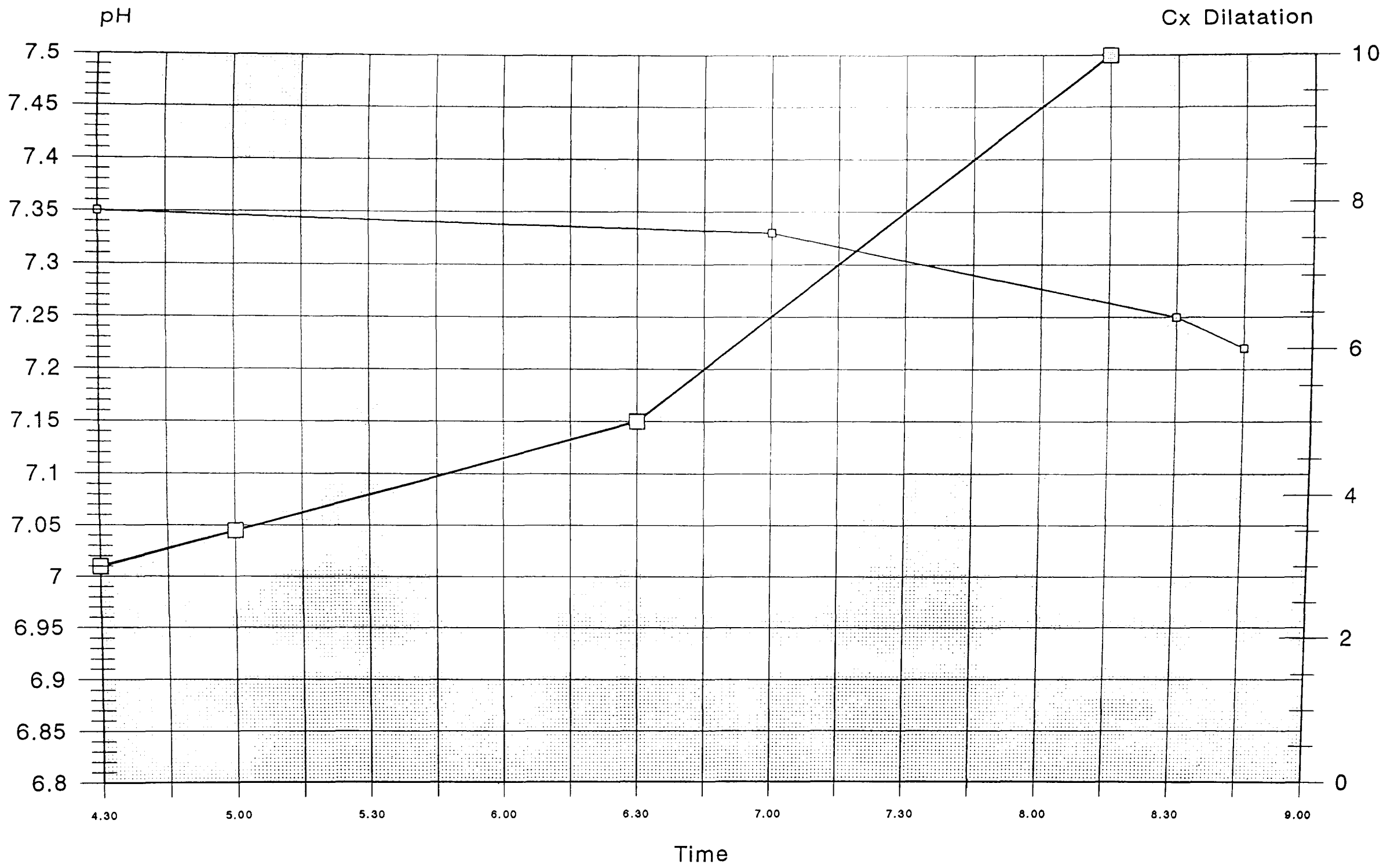
Mrs R.S. is a 23 year old primigravida. She has serious social difficulties and is under psychiatric care because of anxiety and depression. She smokes 20 cigarettes per day. At 38 weeks there was a suspicion of intrauterine growth retardation. An ultrasound scan estimated fetal weight to be 2.4kg, with normal liquor volume. Labour was induced at 41 weeks on the grounds of postmaturity. On admission the fetus was judged to be slightly small for dates (fundal height 36cm). Presentation cephalic 3/5. VE Cx 1cm dilated, 2cm long. A prostin tablet 3mg was given at 15.00hrs on 13.11.91. She was taken to the labour ward at 04.45hrs the following morning with regular contractions. The membranes ruptured spontaneously.

Labour events

04.35 VE Cx 3cm dilated, thick. Station -3. Clear liquor.
05.00 VE Cx 3-4cm dilated. FSE applied.
05.50 Epidural begun.
06.20 Epidural complete. Maternal hypotension; IV fluids increased; oxygen given.
06.35 VE Cx 5cm dilated, thick. Station -2 / -3
07.00 FBS; pH 7.33 BE -2
07.40 Top up.
08.05 Maternal position changed.
08.10 VE Cx fully dilated. Station 0.
08.25 Decision for forceps.
08.47 Male infant delivered by Neville Barnes forceps. Indication; prolonged bradycardia.

Outcome

Birth weight 2.60kg
Apgar 9 & 9
Cord gases pH 7.22 / 7.36 BD(ecf) 4 / 0



○ pH □ Cx Dilatation (cm)

Case 46 180

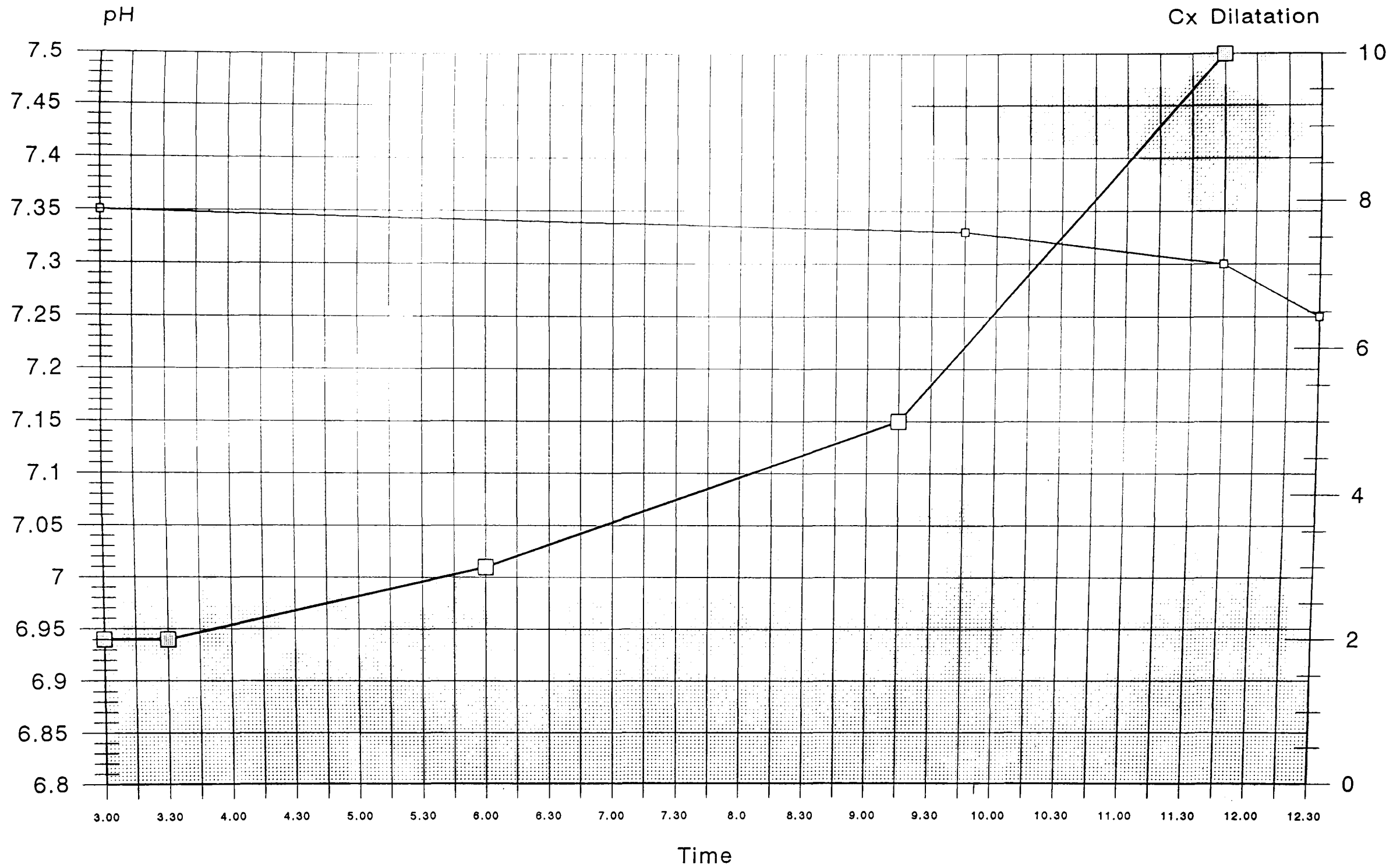
Miss D.R. is a 27 year old primigravida; a fit lady, and a nonsmoker. Her antenatal progress was quite straightforward, and she went into spontaneous labour at 40 weeks. She was admitted on 6/8/90 at 03.05hrs with regular contractions. Membranes had probably ruptured 48 hours previously. Cephalic 2/5. VE Cx 2cm dilated, thick. Station -2. Clear liquor seen.

Labour events

03.30 VE No change. FSE applied.
04.15 Pethidine 100mg, phenergan 25mg im.
04.30 Syntocinon started.
04.45 Vomiting.
06.00 VE Cx 3cm dilated, effaced. Station -2.
07.20 Epidural begun.
07.50 Epidural complete.
09.05 Top up.
09.20 VE Cx 5cm dilated, thick. Station -2. FSE reapplied.
09.50 FBS; pH 7.33 BE -3
10.35 Vomiting.
11.05 Top up.
11.40 Prepared for FBS; Cx found to be fully dilated. Head low and descending well with pushing.
11.50 Pushing begun in earnest.
12.25 Episiotomy.
12.27 Normal delivery of male infant.

Outcome

Birth weight 3.02kg
Apgar 9 & 9
Cord gases pH 7.23 / 7.29 BD(ecf) 2 / 4



□ pH □ Cx Dilatation (cm)

Case 47 1535

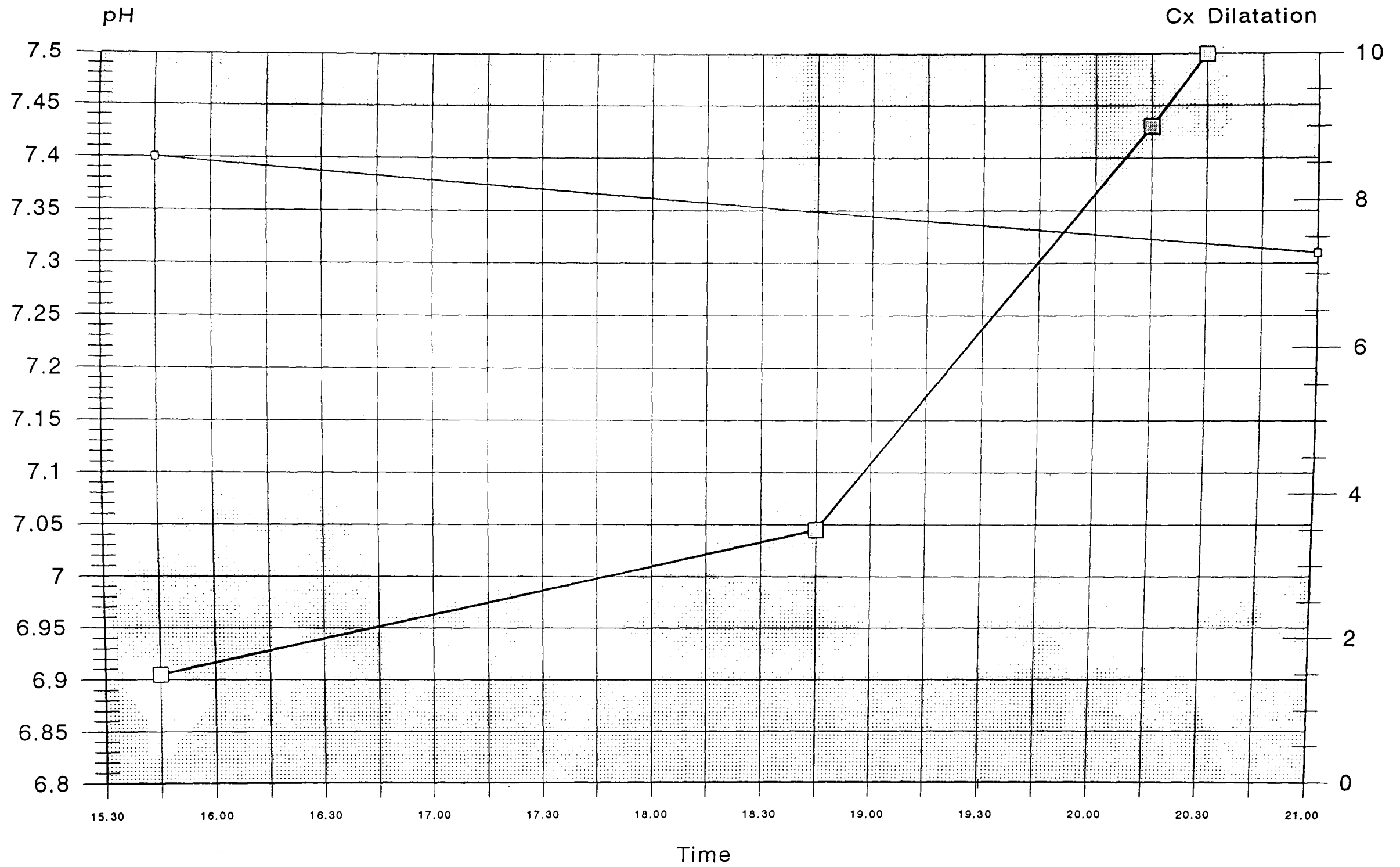
Mrs P.S. is a 30 year old lady expecting her third baby. She is a fit lady and does not smoke. Her two previous children were delivered normally at term, weighing 6lb 11oz and 7lb 6oz respectively. No problems were identified during her current pregnancy. She was admitted on 21.5.91, at 40 weeks gestation with spontaneous rupture of the membranes. Contractions had still not begun 7 hours later and the cervix was just 1-2cm dilated, uneffaced. The liquor was clear. A prostin tablet 3mg was given PV to induce labour. After a further 5 hours no contractions had occurred and VE was unchanged. Labour was therefore induced with intravenous syntocinon, which was commenced at 15.45hrs.

Labour events

18.45 VE Cx 3-4cm dilated, thin. Station -2.
18.55 Pethidine 100mg, phenergan 25mg im.
20.15 VE Cx 9cm dilated. Station -1.
20.35 VE Cx fully dilated. Vertex visible.
20.54 Normal delivery of female infant.

Outcome

Birth weight 3.20kg
Apgar 9 & 9
Cords pH 7.39 BD(ecf) - / 1



Case 48 1337

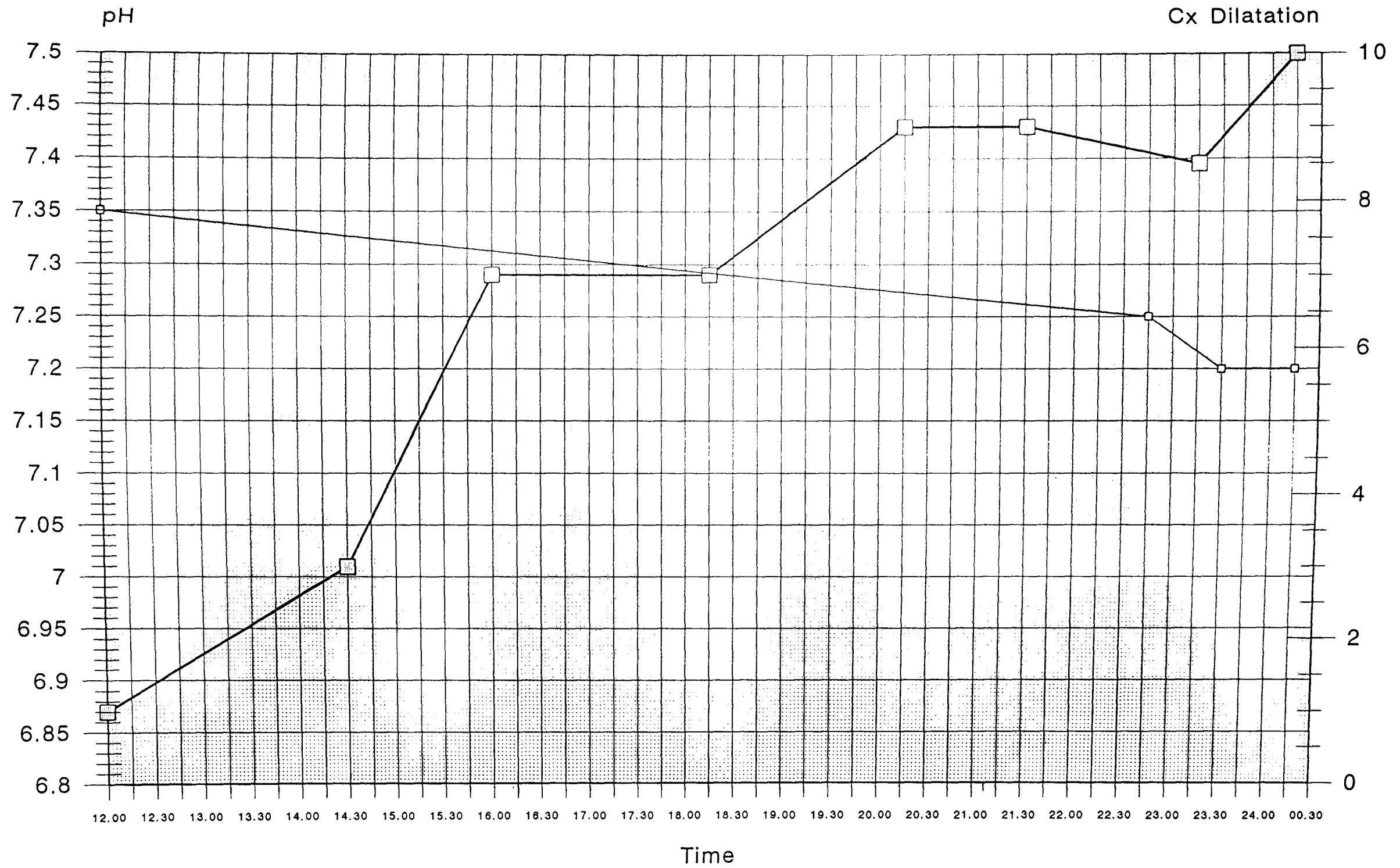
Mrs S.M. is a 25 year old primigravida; a fit lady, and a nonsmoker. Her pregnancy was entirely straightforward and she was admitted in spontaneous labour at 40 weeks on 10/4/91 at 11.30hrs. The uterus was term size; cephalic 3/5. VE (at 12.00hrs) Cx 1cm dilated, effaced and thin. Station -3.

Labour events

- 12.45 Pethidine 100mg phenergan 25mg im.
- 14.30 VE Cx 3cm dilated, thin. Station -2. ARM; clear liquor. FSE applied.
- 15.40 Epidural inserted.
- 15.55 Maternal oxygen given; position changed.
- 16.00 VE Cx 7cm dilated. Station -1. FSE reapplied.
- 18.05 Sat up.
- 18.15 VE No change.
- 18.45 Top up.
- 20.10 Bed pan.
- 20.15 VE Cx 9cm dilated. Station 0. LOT position.
- 20.25 Top up.
- 21.45 VE No change.
- 22.10 Syntocinon started.
- 22.50 Syntocinon stopped, maternal position changed, oxygen given.
- 23.10 VE Cx 8-9cm dilated, thick & oedematous. Station 0. Little caput or moulding.
- 23.25 FBS; pH 7.20 BE -6
Decision for C/S. Indication; secondary arrest of labour, borderline scalp pH.
- 00.15 VE in theatre; Cx now fully dilated and vaginal delivery possible.
- 00.22 Male infant delivered by Neville Barnes forceps.

Outcome

Birth weight 3.13kg
Apgar 6 & 7 (8 at 8 minutes)
Cords pH 7.20 / 7.25 BD(ecf) 4 / 3



Case 49 1542

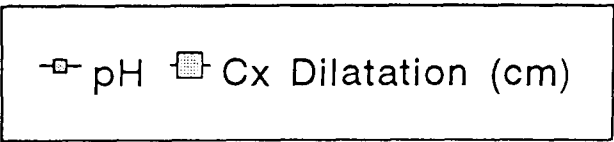
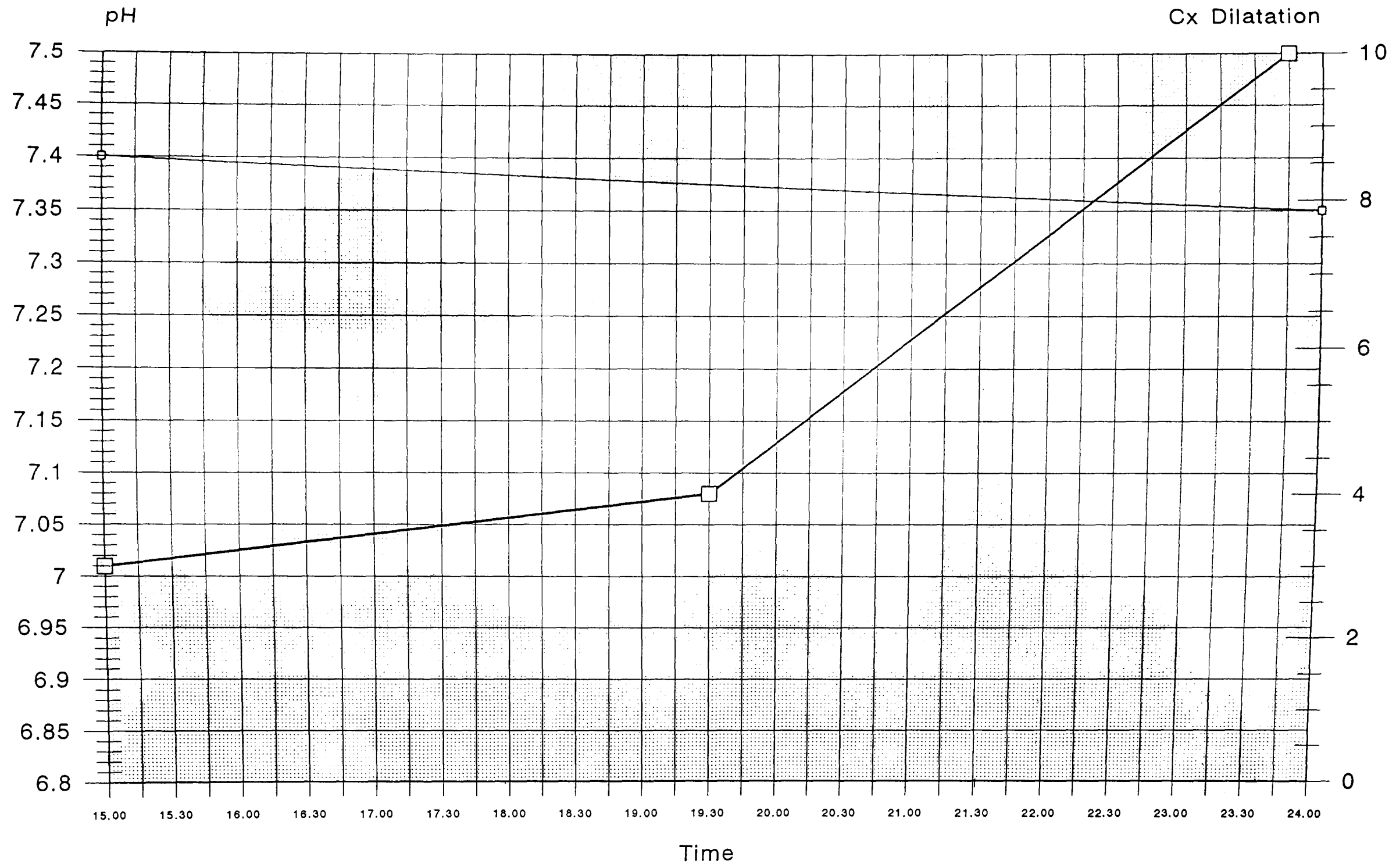
Miss J.M. is a 25 year old lady expecting her fourth child. She smokes 10 cigarettes per day. Her three previous children were delivered normally at term and were well grown. During her current pregnancy there was the suspicion of intrauterine growth retardation and the fetus was monitored with serial ultrasound scans. By 39 weeks there was evidence of asymmetrical growth retardation; estimated fetal weight being 2.1kg. The baby was active and liquor volume normal. It was felt best to induce labour at this stage, but the fetal head was high, the cervix unfavourable, and 12 doses of prostaglandin were required over a period of several days. ARM was eventually performed on 22.5.91 at 15.00hrs; she had reached 40 weeks gestation. The cervix was by now 3cm dilated, 1cm thick. Clear liquor drained in copious amounts.

Labour events

16.40 Syntocinon started.
19.30 VE Cx 4cm dilated, thick. Station -2.
20.20 Failed attempt to insert epidural.
21.10 Pethidine 100mg, phenergan 25mg im.
23.40 VE Cx fully dilated. Vertex visible.
23.50 Pushing begun.
00.09 Normal delivery of female infant.

Outcome

Birth weight 2.48kg
Apgar 8 & 10
Cord gases pH 7.35 / 7.43 BD(ecf) -1 / 3



Case 50 1302

Mrs M.L. is a 31 year old primigravida; fit, and a nonsmoker. There were no antenatal problems, and she laboured spontaneously at 39 weeks. On admission on 3.7.91 the uterus was found to be term size, presentation cephalic 3/5. VE (at 01.50hrs) Cx 2cm dilated, effaced. Station -2.

Labour events

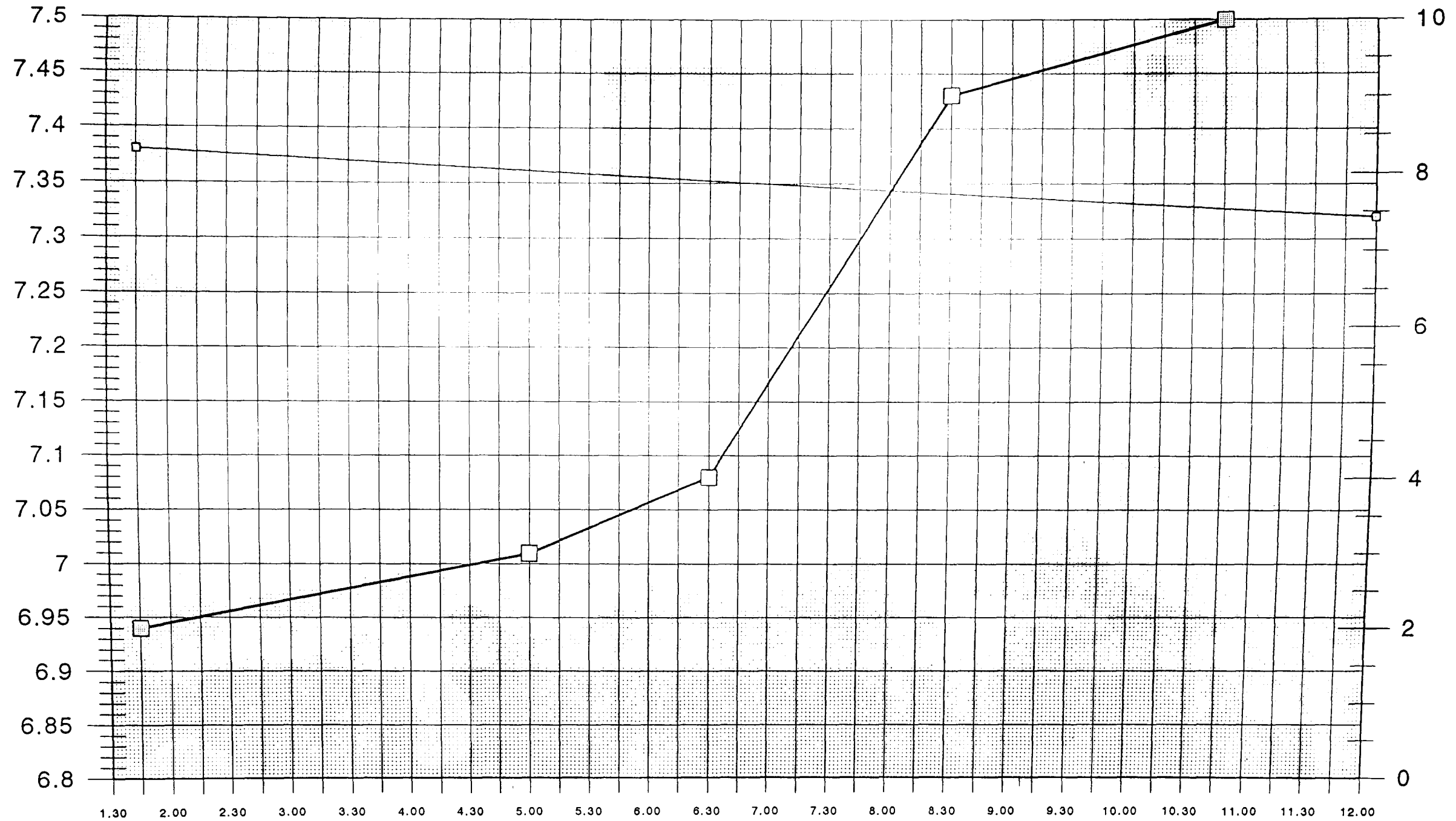
- 04.55 VE Cx 3cm dilated, effaced. Station -1. ARM; meconium stained liquor. FSE applied.
- 05.05 Pethidine 100mg, phenergan 25mg im.
- 06.02 Epidural inserted.
- 06.35 VE Cx 4cm dilated. Station -1. OP position. FSE reapplied.
- 06.50 Syntocinon begun.
- 07.10 Vomited.
- 08.30 VE Cx 9cm dilated. Station 0. OP position.
- 08.45 Top up.
- 09.50 Vomited.
- 10.30 Bedpan.
- 10.40 VE Cx fully dilated. Station +1. OA position.
- 11.00 Vomited.
- 11.05 Pushing begun.
- 12.04 Normal delivery of female infant.

Outcome

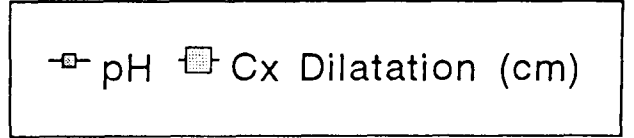
Birth weight 3.30kg
Apgar 9 & 9
Cord gases pH 7.32 / 7.39 BD(ecf) -1 / 0

pH

Cx Dilatation



Time



APPENDIX I

'C' software to calculate the chance expected agreement.

A 'C'-program to calculate the expected chance agreement between 2 reviewers at time t.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>

#define NSCORE    5
#define WWIDTH    3

double calc_e(void);
double E(int i, int j);
double pr(int i, int j, int k, int t);
int f(int i, int k1, int t1, int k2, int t2);
double p1(int k, int t);
double p2(int k, int m, int n);

double prob[NSCORE + 1]=
{
    0, 0.435, 0.438, 0.071, 0.039, 0.017
};

double w[WWIDTH + 1]=
{
    1.00, 0.924, 0.707, 0.383
};

int matrix[NSCORE + 1][NSCORE + 1]=
// Note: this is a doubled matrix to implement as ints.
{
    0, 0, 0, 0, 0, 0,
    0, 4, 3, 0, 0, 0,
    0, 3, 4, 2, 0, 0,
    0, 0, 2, 28, 20, 10,
    0, 0, 0, 20, 50, 40,
    0, 0, 0, 10, 40, 120,
};

int main(int argc, char *argv[])
{
    printf("The expected value is %f\n", calc_e());

    return 0;
}

double calc_e(void)
{
    int i, j;
    double cce= 0, ace= 0;

    for ( i= 1; i <= NSCORE; i++ )
```

```

{
    for ( j= 1; j <= NSCORE; j++ )
    {
        cce+= prob[i] * prob[j] * E(i, j);
    }

    ace+= prob[i] * matrix[i][i];
}

return cce / ace;
}

double E(int i, int j)
{
    int k, t;
    double e= 0;

    for ( k= 1; k <= NSCORE; k++ )
    {
        for ( t= 0; t <= WWIDTH; t++ )
        {
            e+= w[t] * matrix[i][k] * pr(i, j, k, t);
            // To find E[X*X], this line is modified to
            // e+= w[t]*w[t]*matrix[i][k]*matrix[i][k]*pr(i, j, k,
t);
        }
    }

    return e;
}

double pr(int i, int j, int k, int t)
{
    int k1, t1;
    int k2, t2;
    int m, n;
    double pi= 1;

    for ( k1= 1; k1 <= NSCORE; k1++ )
    {
        for ( t1= 1; t1 <= WWIDTH; t1++ )
        {
            for ( m= n= 0, k2= 1; k2 <= NSCORE; k2++ )
            {
                for ( t2= 0; t2 <= WWIDTH; t2++ )
                {
                    m+= f(i, k1, t1, k2, t2);
                    n+= f(i, k, t, k2, t2);
                }
            }

            pi*= p2(k1, m, n);
        }
    }
}

```

```

    return f(i, j, 0, k, t) * p1(k, t) * pi;
}

int f(int i, int k1, int t1, int k2, int t2)
{
    if ( t2 == 0 && k1 == k2 )
        return 1;
    else if ( t2 != 0 && w[t2] * matrix[i][k2] > w[t1] * matrix[i][k1] )
        return 1;
    else
        return 0;
}

double p1(int k, int t)
{
    if ( t != 0 )
        return 1 - (1 - prob[k]) * (1 - prob[k]);
    else
        return 1;
}

double p2(int k, int m, int n)
{
    if ( m < n )
        return (1 - prob[k]) * (1 - prob[k]);
    else
        return 1;
}

```

Appendix J

Tables of results obtained for the Kruskal-Wallis test of inter-agreement.

Tables detailing each reviewers recommendations for caesarean section delivery.

case	review	θ	Reviews which have significantly higher agreement	total
4	A1	46	E1 E2 G1 I2 J2 K1 L1 O1 O2	9
9	A2	73	A1 B2 C1 F1 F2 G1 H1 I2 K2 L1 L2 M2 N2 O1 O2 P2	16
11	A1	68	A2 B2 C1 C2 D1 E1 E2 F2 G1 H1 I1 I2 K1 L1 M2 N2	16
20	A1	64	C2 D1 D2 E1 E2 H1 H2 K1 N1 N2 O2 P2 S1 S2	14
21	A1	68	A2 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H1 H2 I1 I2 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2 S1 S2	32
23	A1	75	A2	1
25	A2	72	B2 D1 D2 E1 E2 G1 H2 J1 J2 K1 L1 L2 M1 N1 N2	15
27	A2	52	A1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 H1 H2 I2 J2 K1 K2 L1 L2 M2 N1 O1 O2 P2 Q1	25
28	A1	63	B2 D1 E1 E2 F2 G1 G2 H1 J2 K2 L2 M2 N2 O1 O2 P2 Q1 S1 S2	19
29	A2	83	D2 L2	2
30	A1	42	A2 D1 D2 E1 H2 J2 L1 L2 M1 M2 O2 P1	12
34	A1	56	A2 C2 D1 D2 E1 E2 G1 H2 I1 I2 K2 L1 L2 M1 M2 O2 P2 S1 S2	19
35	A1	70	A2 D1 D2 E1 E2 G1 H1 I2 J1 J2 K2 L1 N2 O2	14
37	A1	50	C2 D2 H2 L1 L2 O1	6
38	A1	70	C1 C2 D1 D2 E1 E2 I1 I2 N2	9
39	A1	89	B1 B2 C1 D1 D2 E1 E2 F1 F2 G1 G2 H2 I1 I2 J2 K1 K2 L1 L2 M1 M2 N2 O1 P1 P2 S1 S2	27
39	A2	85	B1 B2 C1 D1 D2 E1 E2 H2 I1 I2 J2 K1 L1 L2 O1 P1 P2 S1 S2	19
50	A2	64	A1 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H2 I1 I2 J1 J2 K1 K2 M1 M2 N1 O2 Q1 S1 S2	27

Table J.1: Cases where at least 1 review obtained higher agreements than expert A.

Summary. The majority of other reviews obtained significantly higher agreement than expert A in 6 cases (7 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
2	B1	28	E2 O2	2
3	B1	35	C1 D2 E2 H2 I1 J2 K2 O2	8
4	B2	25	A2 B1 C1 C2 D1 D2 E1 E2 G1 I1 I2 J2 K1 L1 L2 M2 N1 O1 O2 S1 S2	21
5	B1	50	A2 D1 I2 L1 O2	5
5	B2	40	A2 C1 C2 D1 F2 I1 I2 J2 L1 L2 N2 O2 S1 S2	14
8	B2	67	E2 I1	2
11	B1	68	A2 B2 C1 C2 D1 E1 E2 F2 G1 H1 I1 I2 K1 L1 M2 N2	16
12	B1	35	A2 E1 G1 G2 I2 K1 L1 L2 O1 O2 P2 S1 S2	13
12	B2	23	A1 A2 C2 D1 E1 F2 G1 G2 H1 H2 I1 I2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P2 S1 S2	25
13	B2	40	A1 A2 C1 C2 E2 F1 G1 G2 H1 H2 I1 I2 J1 K1 K2 L1 L2 M1 M2 O1 O2 P1 P2 S1 S2	25
14	B1	39	A1 C1 D1 D2 E1 E2 F1 F2 G1 G2 H2 I1 I2 J1 J2 K1 K2 L1 L2 M1 M2 N2 O1 O2 P1 P2 S1 S2	28
14	B2	44	A1 C1 D1 D2 E1 E2 F1 F2 G1 G2 H2 I1 I2 J1 J2 K1 K2 L1 L2 M1 M2 N2 O1 O2 P1 P2	26
15	B1	11	A1 A2 C1 D1 D2 E1 E2 F1 G1 G2 I1 I2 J1 J2 K2 L2 M1 M2 N1 N2 O1 O2 P1 P2 Q1 Q2 S1 S2	28
15	B2	23	A1 A2 C1 D1 D2 E1 E2 F1 G1 G2 I1 I2 J1 J2 K2 L2 M1 M2 N1 N2 O1 O2 P1 P2 Q1 Q2 S1 S2	28
16	B1	22	A1 A2 C1 C2 D1 E1 E2 F2 G1 G2 H1 H2 I1 I2 J1 K1 L1 L2 M1 M2 N2 O1 O2 P1 S1 S2	26
16	B2	31	A1 A2 E1 F2 H1 H2 M2 S1 S2	9
18	B1	41	A1 A2 C1 C2 D1 D2 E1 E2 F1 G1 G2 H1 H2 I1 I2 J1 K1 K2 L1 L2 M2 O1 O2 P1 P2 S1 S2	27
18	B2	43	A1 A2 C1 C2 D1 D2 E1 E2 F1 G1 G2 H1 H2 I1 I2 J1 K1 K2 L1 O1 O2 P1 P2	23
20	B1	71	D2 P2 S1 S2	4
22	B1	11	A1 A2 C1 D1 D2 E1 E2 F1 G1 G2 I1 I2 J1 J2 K2 L2 M1 M2 N1 N2 O1 O2 P1 P2 Q1 Q2 S1 S2	28
22	B2	23	A1 A2 C1 D1 D2 E1 E2 F1 G1 G2 I1 I2 J1 J2 K2 L2 M1 M2 N1 N2 O1 O2 P1 P2 Q1 Q2 S1 S2	28
24	B1	31	A1 A2 C1 C2 D1 E1 F1 F2 G1 G2 I1 I2 J1 J2 K1 K2 L1 L2 M2 N1 N2 O2 P2	23
26	B1	74	E2	1
26	B2	74	E2 G1 H1 I2 P2	5
27	B1	67	A1 C1 C2 D1 D2 E2 F1 F2 G1 H1 H2 I2 J2 K1 K2 L1 L2 M2 O1 O2 P2 Q1	22
28	B1	59	D1 E1 E2 F2 G1 G2 I1 J2 K2 L2 M2 N2 O1 O2 P2 Q1	16
29	B1	78	D2 L2 N2	3
29	B2	80	D2 E1 E2 H2 I1 J2 L2 N2 O1	9
30	B1	31	A2 B2 C2 D1 D2 E1 H1 H2 J2 K1 L1 L2 M1 M2 N2 O2 P1 Q2	18
33	B1	58	A1 A2 C1 D1 D2 E1 F1 F2 G1 G2 H2 J2 K1 K2 L1 L2 M1 M2 N1 N2 P1 P2 Q2	23
33	B2	44	A1 A2 C1 D1 D2 E1 F1 F2 G1 G2 H2 J2 K1 K2 L1 L2 M1 M2 N1 N2 P1 P2 Q2 S1 S2	25
34	B1	57	A2 C2 D1 G1 H2 I1 I2 K2 L1 O2 S1	11
34	B2	57	A2 C2 D1 G1 H2 I1 I2 K2 L1 O2 S1	11
35	B1	79	A2 D1 D2 H1 I2 J1 J2 K2 L1 N2 O2	11
35	B2	65	A2 D1 D2 E1 E2 G1 H1 H2 I2 J1 J2 K1 K2 L1 N1 N2 O1 O2 S2	19
36	B2	44	A1 A2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H1 H2 I1 I2 J2 K1 K2 L1 L2 M1 N1 N2 O1 O2 P1 S1 S2	29
37	B1	48	C2 D2 G1 H2 L1 L2 N1 O1	8
38	B1	61	C1 C2 D2 E1 I1 I2 N2	7
40	B1	22	A2 C2 D1 E1 F1 F2 G1 I2 J2 K1 L1 L2 M1 M2 N1 O2	16
41	B1	44	B2 C1 D1 D2 E1 E2 I2 J2 K1 L1 N1 O1 O2	13
46	B1	62	A2 C1 C2 D1 D2 F1 F2 G1 G2 H1 H2 I1 I2 J2 K1 K2 L1 L2 M1 M2 N2 O1 O2 P1 P2 Q1 S1 S2	28
46	B2	72	D1 D2 G1 G2 H2 I1 J2 K2 L1 M2 P2 S1 S2	13
49	B1	73	A1 A2 C1 C2 D1 D2 E1 E2 G1 G2 H1 H2 I2 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 Q1 S1 S2	29
49	B2	87	A2 C1 C2 D1 E1 G1 H1 J2 K2 L1 L2 M1 N1 N2 O1 O2 P1 S1 S2	19

Table J.2: Cases where at least 1 review obtained higher agreements than expert B.

Summary. The majority of other reviews obtained significantly higher agreement than expert B in 16 cases (22 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
2	C1	19	A1 A2 C2 D1 D2 E2 G1 G2 I1 I2 J1 K1 L1 L2 M1 M2 O2 P1 P2 S1 S2	21
8	C2	65	D2 E2 I1	3
25	C1	61	A1 B2 D1 D2 E1 E2 F2 G1 G2 H2 I2 J1 J2 K1 L1 L2 M1 M2 N1 N2 O2 Q1 Q2 S1 S2	25
25	C2	61	A1 B2 D1 D2 E1 E2 F2 G1 G2 H2 I2 J1 J2 K1 L1 L2 M1 M2 N1 N2 O2 Q1 Q2 S1 S2	25
26	C1	70	A2 D1 D2 E2 F1 F2 G1 H1 I2 M2 N1 N2 P1 P2	14
28	C1	0	A2 B2 C2 D1 E1 E2 F1 F2 G1 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 L2 M1 M2 N2 O1 O2 P1 P2 Q1 S1 S2	30
33	C2	49	A1 A2 C1 D1 D2 E1 F1 F2 G1 G2 H2 J2 K1 K2 L1 L2 M1 M2 N1 N2 P1 P2 Q2	23
39	C2	84	B1 B2 C1 D1 D2 E1 E2 F1 F2 G1 G2 H1 H2 I1 I2 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 P1 P2 S1 S2	29
47	C2	81	A1 A2 B1 B2 C1 D1 D2 F1 F2 G2 H1 I1 I2 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2	28

Table J.3: Cases where at least 1 review obtained higher agreements than expert C.

Summary. The majority of other reviews obtained significantly higher agreement than expert C in 6 cases (7 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
6	D1	90	A2 B1 B2 C1 C2 D2 E1 E2 F1 F2 G1 G2 H1 I1 I2 K1 K2 L1 L2 N1 N2 O1 P1 P2	24
7	D1	37	A1 A2 B1 B2 C1 D2 E2 F1 F2 H2 I1 I2 J2 K1 M1 P1 P2 Q1 S1 S2	20
28	D2	63	D1 E1 F2 G2 I1 J2 K2 M2 N2 O1 O2 P2	12
31	D2	97	A1 C1 C2 D1 F1 F2 G1 G2 H2 I2 J1 J2 K1 K2 L1 L2 M1 M2 N2 O1 O2 P1 P2	23
32	D2	28	C2 H2 J2 L2	4
37	D1	50	C2 H2 L2 O1	4

Table J.4: Cases where at least 1 review obtained higher agreements than expert D.

Summary. The majority of other reviews obtained significantly higher agreement than expert D in 3 cases (3 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
7	E1	40	C1 D2 E2 P1 S1 S2	6
12	E2	27	A1 A2 E1 G1 G2 H1 H2 I2 K1 K2 L1 L2 O1 O2 P2 S1 S2	17
31	E2	97	A1 C1 C2 D1 F1 F2 G1 G2 H2 I2 J1 J2 K1 K2 L1 L2 M1 M2 N2 O1 O2 P1 P2	23
37	E1	50	B2 C2 D2 F1 G1 H2 J1 K1 L1 L2 N1 O1 P1 P2	14
37	E2	51	C2 D2 G1 H2 L1 L2 N1 O1	8
46	E1	74	G2	1
47	E2	91	C1 F1 H1 I1 L2 N1	6

Table J.5: Cases where at least 1 review obtained higher agreements than expert E.

Summary. The majority of other reviews obtained significantly higher agreement than expert E in 1 case (1 review).

case	review	θ	Reviews which have significantly higher agreement	total
1	F1	54	B2 C2 E1 H1 I2 K1 N1 Q1	8
3	F2	16	A1 A2 C1 C2 D1 D2 E1 E2 G2 H2 I1 I2 J2 K2 L1 L2 N1 N2 O1 O2 P1 S1 S2	23
8	F1	66	A1 A2 B1 C1 D2 E2 F2 G1 H1 H2 I1 J2 K1 K2 L2 M1 M2 N1 O1 P1 P2 Q2	22
10	F1	10	A1 A2 B1 B2 C1 C2 D1 D2 E1 E2 F2 G1 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 L2 N2 O1 O2 P1 P2 Q1 S1 S2	31
12	F1	33	A1 A2 E1 G1 G2 H1 H2 I2 K1 K2 L1 L2 O1 O2 P2 S1 S2	17
16	F1	35	S1 S2	2
19	F1	39	I1 I2 J1 K2 L2 M2 P1	7
23	F1	73	A2 I2	2
23	F2	73	A2 I2	2
41	F1	46	B2 C1 D1 D2 E1 E2 I2 J2 K1 K2 L1 N1 O1 O2	14
42	F1	46	C1 C2 D2 H1 I1 I2 J2 N2 P1 P2	10
42	F2	48	C1 C2 D2 I1 J2 N2 P1 P2	8
43	F1	40	A1 A2 B1 B2 C1 C2 D1 D2 E1 E2 F2 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 L2 M2 O1 O2 P2	26
49	F1	86	A2 C1 C2 D1 E1 G1 G2 H1 J2 K2 L1 L2 M1 N1 N2 O1 O2 P1 S1 S2	20
49	F2	86	A2 C1 C2 D1 E1 G1 G2 H1 I2 J2 K1 K2 L1 L2 M1 N1 N2 O1 O2 P1 P2 S1 S2	23

Table J.6: Cases where at least 1 review obtained higher agreements than expert F.

Summary. The majority of other reviews obtained significantly higher agreement than expert F in 5 cases (6 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
7	G2	52	A1 A2 B1 B2 C1 C2 D2 E2 F1 F2 H2 I1 I2 J2 K1 K2 M1 M2 N1 N2 P1 P2 Q1 Q2 S1 S2	26
8	G2	52	A1 A2 B1 C1 D1 D2 E1 E2 F2 G1 H1 H2 I1 I2 J2 K1 K2 L2 M1 M2 N1 O1 O2 P1 P2 Q2	8
26	G2	72	D2 E2 F1 G1 H1 I2 N2 P2	2
29	G2	83	D2 L2	4
34	G2	59	D1 G1 K2 S1	3
40	G2	26	C2 L1 O2	10
43	G1	88	A2 D1 D2 E1 H2 J2 L2 M2 O1 O2	13
45	G1	60	A2 D1 D2 E1 E2 I2 K1 K2 L1 L2 P2 S1 S2	1
45	G2	57	E2	

Table J.7: Cases where at least 1 review obtained higher agreements than expert G.

Summary. The majority of other reviews obtained significantly higher agreement than expert G in 2 cases (2 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
2	H1	5	A1 A2 B2 C2 D1 D2 E1 E2 F1 F2 G1 G2 I1 I2 J1 K1 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2	28
2	H2	5	A1 A2 B2 C2 D1 D2 E1 E2 F1 F2 G1 G2 I1 I2 J1 K1 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2	28
4	H1	29	C1 C2 D1 D2 E1 E2 G1 I1 I2 J2 K1 L1 L2 M2 O1 O2 S1 S2	18
4	H2	29	C1 C2 D1 D2 E1 E2 G1 I1 I2 J2 K1 L1 L2 M2 O1 O2 S1 S2	18
6	H2	89	A2 B1 B2 C1 C2 D2 E1 E2 F1 F2 G1 G2 H1 I1 I2 K1 K2 L1 L2 N1 N2 O1 P1 P2	24
15	H1	4	A1 A2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 I1 I2 J1 J2 K1 K2 L2 M1 M2 N1 N2 O1 O2 P1 P2 Q1 Q2 S1 S2	31
15	H2	35	A1 A2 C1 D1 D2 E1 E2 F1 G1 G2 I1 I2 J1 J2 K2 L2 M1 M2 N1 N2 O1 O2 P1 P2 Q1 Q2 S1 S2	28
17	H1	28	A1 A2 C1 C2 D1 F2 G2 H2 I1 I2 J1 K1 L1 L2 N1 N2 O2	17
19	H1	34	K2 L2	2
22	H1	4	A1 A2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 I1 I2 J1 J2 K1 K2 L2 M1 M2 N1 N2 O1 O2 P1 P2 Q1 Q2 S1 S2	31
22	H2	35	A1 A2 C1 D1 D2 E1 E2 F1 G1 G2 I1 I2 J1 J2 K2 L2 M1 M2 N1 N2 O1 O2 P1 P2 Q1 Q2 S1 S2	28
24	H1	46	A2 G1 J2 L2 N2 P2	6
24	H2	23	A1 A2 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 I1 I2 J1 J2 K1 K2 L1 L2 M2 N1 N2 O2 P2	26
25	H1	77	B2 D1 E1 E2 G1 H2 J1 K1 L1 L2 N1 N2	12
37	H1	41	A2 B2 C1 C2 D2 F1 F2 G1 G2 H2 I1 I2 J1 J2 K1 K2 L1 L2 N1 N2 O1 O2 P1 P2 S1 S2	26
38	H1	62	C1 C2 D2 E1 I1 I2 N2	7
38	H2	65	C1 C2 D1 D2 E1 E2 G1 I1 I2 N2	10
41	H1	37	A2 B2 C1 C2 D1 D2 E1 E2 F2 G1 G2 I1 I2 J1 J2 K1 K2 L1 N1 N2 O1 O2 Q2 S1 S2	25
44	H1	31	A1 A2 B1 B2 C1 C2 D1 D2 E1 E2 F2 G1 G2 I1 I2 J1 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P2 Q1	29
50	H1	83	A1 B2 E1 I1 J2 O2 S1 S2	8

Table J.8: Cases where at least 1 review obtained higher agreements than expert H.

Summary. The majority of other reviews obtained significantly higher agreement than expert H in 9 cases (13 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
20	I1	72	C2 D2 E2 H2 P2 S1 S2	7
20	I2	69	C2 D2 E2 H2 P2 S1 S2	7
27	I1	71	A1 C1 D1 D2 F1 F2 G1 H1 K1 L1 L2 O2	12
31	I1	98	A1 C1 D1 G1 H2 L1 N2 O1	8
48	I2	54	C2	1
49	I1	86	A2 C1 C2 D1 E1 G1 G2 H1 I2 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2	24

Table J.9: Cases where at least 1 review obtained higher agreements than expert I.

Summary. The majority of other reviews obtained significantly higher agreement than expert I in 1 case (1 review).

case	review	θ	Reviews which have significantly higher agreement	total
3	J1	24	A1 A2 C1 C2 D1 D2 E1 E2 G2 H2 I1 I2 J2 K2 L1 L2 N1 N2 O1 O2 P1 S1 S2	23
6	J1	94	A2 D2 H1 O1	4
6	J2	88	A2 B1 B2 C1 C2 D2 E1 E2 F1 F2 G1 G2 H1 I1 I2 K1 K2 L1 L2 N1 N2 O1 P1 P2	24
18	J2	46	A1 A2 C1 C2 D1 D2 E1 E2 F1 G1 G2 H1 H2 I1 I2 J1 K1 K2 L1 O1 O2 P1 P2	23
20	J2	69	D2 P2 S1 S2	4
21	J1	58	A2 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H1 H2 I1 I2 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2	32
23	J2	70	A2 E2 G1 I1 I2 M2	6
26	J1	61	A2 D1 D2 E1 E2 F1 F2 G1 H1 H2 I1 I2 K2 L2 M1 M2 N1 N2 O2 P1 P2 Q1 S1 S2	24
33	J1	62	A1 A2 C1 D1 D2 E1 F1 F2 G1 G2 H2 J2 K1 K2 L1 L2 M1 M2 N1 N2 P1 P2 Q2	23
34	J1	42	A2 C1 C2 D1 D2 E1 E2 F1 F2 G1 H1 H2 I1 I2 K1 K2 L1 L2 M1 M2 N1 N2 O2 P1 P2 S1 S2	27
39	J1	84	B1 B2 C1 D1 D2 E1 E2 F1 F2 G1 G2 H2 I1 I2 J2 K1 K2 L1 L2 M1 M2 N2 O1 P1 P2 S1 S2	27
46	J1	61	C1 D1 D2 F1 F2 G1 G2 H1 H2 I1 I2 J2 K1 K2 L1 L2 M1 M2 O1 O2 P1 P2 Q1 S1 S2	25
47	J1	81	A1 A2 B1 B2 C1 D1 D2 F1 F2 G2 H1 H2 I1 I2 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2	29
49	J1	86	A2 C1 C2 D1 E1 G1 G2 H1 J2 K2 L1 L2 M1 N1 N2 O1 O2 P1 S1 S2	20

Table J.10: Cases where at least 1 review obtained higher agreements than expert J.

Summary. The majority of other reviews obtained significantly higher agreement than expert J in 11 cases (11 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
1	K2	56	K1	1
3	K1	8	A1 A2 B2 C1 C2 D1 D2 E1 E2 F1 G1 G2 H2 I1 I2 J2 K2 L1 L2 N1 N2 O1 O2 P1 P2 S1 S2	27
25	K2	61	A1 B2 D1 D2 E1 E2 F2 G1 G2 H2 I2 J1 J2 K1 L1 L2 M1 M2 N1 N2 O2 Q1 Q2 S1 S2	25
29	K2	78	A1 C1 C2 D2 E1 E2 H2 I1 I2 J2 L2 N2 O1 P1	14
38	K2	70	C1 C2 D1 D2 E1 E2 I1 I2 N2	9

Table J.11: Cases where at least 1 review obtained higher agreements than expert K.

Summary. The majority of other reviews obtained significantly higher agreement than expert K in 2 cases (2 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
7	L1	41	C1 D2 E2 P1 S1 S2	6
7	L2	28	A1 A2 B1 B2 C1 D2 E2 F1 F2 H2 I1 I2 J2 K1 K2 M1 N1 N2 P1 P2 Q1 Q2 S1 S2	24
8	L1	72	A2 D2 E2 I1 P1	5
15	L1	58	A1 D2 E1 E2 G1 G2 L2 M1 M2 N1 Q1 Q2	12
20	L2	72	D2 P2 S1 S2	4
22	L1	58	A1 D2 E1 E2 G1 G2 L2 M1 M2 N1 Q1 Q2	12
26	L1	69	A2 D1 D2 E2 F1 F2 G1 H1 I1 I2 M2 N1 N2 P1 P2	15
28	L1	67	I1 O1	2
48	L2	19	A1 A2 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H1 H2 I1 J1 J2 K1 K2 L1 N1 N2 O2 P1 P2 Q1 Q2	29
50	L1	78	A1 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H2 I1 I2 J1 J2 K1 K2 M1 M2 O2 S1 S2	25
50	L2	85	B2 J2 S1 S2	4

Table J.12: Cases where at least 1 review obtained higher agreements than expert L.

Summary. The majority of other reviews obtained significantly higher agreement than expert L in 3 cases (3 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
1	M1	54	B2 C2 E1 H1 I2 K1 N1	7
3	M1	42	C1 D2 E2 H2 I1 I2 J2 K2 O1 O2	10
3	M2	42	C1 D2 E2 H2 I1 I2 J2 K2 O1 O2	10
6	M1	89	A2 B1 B2 C1 C2 D2 E1 E2 F1 F2 G1 G2 H1 I1 I2 K1 K2 L1 L2 N1 N2 O1 P1 P2	24
6	M2	32	A2 B1 B2 C1 C2 D2 E1 E2 F1 F2 G1 G2 H1 I1 I2 K1 K2 L1 L2 N1 N2 O1 P1 P2 Q1 Q2 S1 S2	28
10	M1	12	A2 B1 B2 C1 C2 D1 D2 E1 E2 F2 G1 G2 H1 H2 I1 I2 J1 K1 K2 L1 L2 N2 O1 O2 P1 P2 Q1 S1 S2	29
10	M2	25	A2 B1 C1 C2 D2 E2 G1 G2 H2 I1 I2 J1 K1 L1 L2 O1 P1 S1 S2	19
17	M1	36	A1 A2 C1 C2 D1 F2 G2 H2 I1 I2 J1 K1 L1 L2 N1 N2 O2	17
24	M1	44	A1 A2 C2 D1 E1 F1 F2 G1 I1 I2 J1 J2 K1 K2 L1 L2 M2 N1 N2 O2 P2	21
27	M1	42	A1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 H1 H2 I2 J1 J2 K1 K2 L1 L2 M2 N1 N2 O1 O2 P2 Q1 S1 S2	29
28	M1	68	I1	1
32	M1	29	A2 C2 H2 J1 J2 L2	6
35	M1	69	A2 D1 D2 E1 E2 G1 H1 I2 J1 J2 K2 L1 N2 O2	14
41	M1	45	A2 B2 C1 D1 D2 E1 E2 G1 I1 I2 J1 J2 K1 K2 L1 N1 O1 O2	18
41	M2	48	B2 C1 D1 D2 E1 E2 I2 J2 L1 N1 O1 O2	12
42	M1	45	C1 C2 D2 H1 I1 I2 J2 N2 O1 P1 P2	11
42	M2	48	D2 N2	2
43	M1	54	A1 A2 C2 D1 D2 E1 E2 F2 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 L2 M2 O1 O2 P2	23
45	M1	51	A1 A2 B2 C1 C2 D1 D2 E1 E2 F2 H1 H2 I1 I2 K1 K2 L1 L2 N1 N2 O1 O2 P2 S1 S2	25
48	M1	56	C2 K2	2
48	M2	56	B2 C1 C2 G2 K2 L1 P1	7

Table J.13: Cases where at least 1 review obtained higher agreements than expert M.

Summary. The majority of other reviews obtained significantly higher agreement than expert M in 7 cases (9 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
8	N2	52	A1 A2 B1 C1 D1 D2 E1 E2 F2 G1 H1 H2 I1 I2 J2 K1 K2 L2 M1 M2 N1 O1 O2 P1 P2 Q2	26
13	N1	39	A1 A2 C1 C2 E2 F1 G1 G2 H1 H2 I1 I2 J1 K1 K2 L1 L2 M1 M2 O1 O2 P1 P2 S1 S2	25
16	N1	33	S1 S2	2
28	N1	60	B2 D1 E1 E2 F2 G1 G2 H2 I1 J2 K2 L2 M2 N2 O1 O2 P2 Q1 S1 S2	20
29	N1	82	D2 L2	2
38	N1	65	C1 C2 D1 D2 E1 E2 I1 I2 N2	9
40	N2	23	A1 A2 C1 C2 D1 D2 E1 F1 F2 G1 I1 I2 J1 J2 K1 K2 L1 L2 M1 M2 N1 O2 P1 P2	24
43	N2	64	A2 D1 D2 E1 H1 H2 I2 J2 K1 L2 M2 O1 O2	13
50	N2	84	B2 E1 J2 O2 S1 S2	6

Table J.14: Cases where at least 1 review obtained higher agreements than expert N.

Summary. The majority of other reviews obtained significantly higher agreement than expert N in 4 cases (4 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
6	O2	89	A2 B1 B2 C1 C2 D2 E1 E2 F1 F2 G1 G2 H1 I1 I2 K1 K2 L1 L2 N1 N2 O1 P1 P2	24
7	O1	43	D2 E2 S1 S2	4
23	O1	72	A2 I2	2
24	O1	23	A1 A2 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H1 I2 J1 J2 K1 K2 L1 L2 M2 N1 N2 O2 P2	26
25	O1	70	B2 D1 D2 E1 E2 G1 H2 J1 J2 K1 L1 L2 M1 N1 N2	15
26	O1	70	A2 D1 D2 E2 F1 F2 G1 H1 I1 I2 M2 N1 N2 P1 P2	15
34	O1	64	A2 C2 D1 G1 H2 I1 I2 K2 L1 O2 S1	11
39	O2	84	B1 B2 C1 D1 D2 E1 E2 F1 H2 I1 I2 J2 K1 L1 L2 O1 P1 P2 S1 S2	20
40	O1	23	A2 C2 D1 E1 F1 G1 J2 K1 L1 L2 M1 N1 O2	13
48	O1	19	A1 A2 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H1 H2 I1 J1 J2 K1 K2 L1 N1 N2 O2 P1 P2 Q1 Q2	29
50	O1	84	A1 B2 C1 D1 D2 E1 E2 H2 I1 J1 J2 K2 M1 O2 S1 S2	16

Table J.15: Cases where at least 1 review obtained higher agreements than expert O.

Summary. The majority of other reviews obtained significantly higher agreement than expert O in 4 cases (4 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
20	P1	70	D2	1
24	P1	24	A1 A2 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H1 I2 J1 J2 K1 K2 L1 L2 M2 N1 N2 O2 P2	26
25	P1	72	B2 D1 D2 E1 E2 G1 H2 J1 J2 K1 L1 L2 M1 N1 N2	15
27	P1	55	A1 C1 C2 D1 D2 E2 F1 F2 G1 H1 H2 I2 J2 K1 K2 L1 L2 M2 N1 O1 O2 P2 Q1	23
41	P1	47	B2 C1 D1 D2 E1 E2 I2 J2 K1 L1 N1 O1 O2	13
41	P2	46	B2 C1 D1 D2 E1 E2 I2 J2 K1 K2 L1 N1 O1 O2	14
44	P1	28	A1 A2 B2 C1 C2 D1 E1 F2 G2 H1 I2 J2 K1 K2 L1 L2 N2 P2 Q1	19
45	P1	58	A2 B2 C1 D1 D2 E1 E2 F2 H1 H2 I2 K1 K2 L1 L2 O2 P2 S1 S2	19
50	P1	74	A1 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H2 I1 I2 J1 J2 K1 K2 M1 M2 N1 O2 S1 S2	26
50	P2	83	A1 B1 B2 C1 D1 D2 E1 E2 F1 H2 I1 I2 J1 J2 K1 K2 M1 M2 O2 S1 S2	21

Table J.16: Cases where at least 1 review obtained higher agreements than expert P.

Summary. The majority of other reviews obtained significantly higher agreement than expert P in 5 cases (6 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
2	Q1	22	A1 C2 E2 I2 L1 L2 O2	
3	Q1	15	A1 A2 C1 C2 D1 D2 E1 E2 G2 H2 I1 I2 J2 K2 L1 L2 N1 N2 O1 O2 P1 S1 S2	7
4	Q1	44	J2	23
5	Q1	8	A1 A2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H2 I1 I2 J1 J2 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2	1
5	Q2	35	A2 C1 C2 D1 D2 F1 F2 G1 G2 I1 I2 J1 J2 L1 L2 N2 O2 P2 S1 S2	30
8	Q1	38	A1 A2 B1 C1 D1 D2 E1 E2 F2 G1 H1 H2 I1 I2 J1 J2 K1 K2 L2 M1 M2 N1 O1 O2 P1 P2 Q2	20
9	Q2	73	A1 C1 I2 K2 N2 O1 O2 P2	27
11	Q1	67	A2 B2 C1 C2 D1 E1 E2 F2 G1 H1 H2 I1 I2 K1 L1 M2 N2	8
12	Q1	26	A1 A2 E1 G1 G2 H1 H2 I2 K1 K2 L1 L2 O1 O2 P2 S1 S2	17
12	Q2	26	A1 A2 D1 E1 G1 G2 H1 H2 I2 K1 K2 L1 L2 O1 O2 P2 S1 S2	17
13	Q1	21	A1 A2 C1 C2 D2 E1 E2 F1 F2 G1 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 L2 M1 M2 N2 O1 O2 P1 P2 S1 S2	18
13	Q2	21	A1 A2 C1 C2 D2 E1 E2 F1 F2 G1 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 L2 M1 M2 N2 O1 O2 P1 P2 S1 S2	30
14	Q1	16	A1 A2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 Q2 S1 S2	30
16	Q1	17	A1 A2 C1 C2 D1 E1 E2 F2 G1 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 L2 M1 M2 N2 O1 O2 P1 P2 S1 S2	33
16	Q2	19	A1 A2 C1 C2 D1 E1 E2 F2 G1 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 L2 M1 M2 N2 O1 O2 P1 P2 S1 S2	29
17	Q1	33	A1 C1 D1 G2 I1 J1 K1 L1 L2 N1 N2	29
18	Q1	34	A1 A2 C1 C2 D1 D2 E1 E2 F1 G1 G2 H1 H2 I1 I2 J1 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2	11
18	Q2	30	A1 A2 C1 C2 D1 D2 E1 E2 F1 G1 G2 H1 H2 I1 I2 J1 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2	30
20	Q1	67	D2 P2	2
21	Q1	88	A2 C2 D1 D2 E1 E2 H2 J2 K1 O1 O2 P1 P2 S1 S2	15
21	Q2	83	A2 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G2 H2 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2	28
23	Q1	11	A1 A2 B1 B2 C1 C2 D1 D2 E1 E2 G1 G2 H1 H2 I1 I2 J1 K1 K2 L1 L2 M1 M2 N1 N2 O2 P1 P2 S1 S2	30
23	Q2	3	A1 A2 B1 B2 C1 C2 D1 D2 E1 E2 F2 G1 G2 H1 H2 I1 I2 J1 K1 K2 L1 L2 M1 M2 N1 N2 O2 P1 P2 S1 S2	31
26	Q2	26	A1 A2 B1 C2 D1 D2 E1 E2 F1 F2 G1 H1 H2 I1 I2 J2 K1 K2 L2 M1 M2 N1 N2 O2 P1 P2 Q1 S1 S2	29
27	Q2	53	A1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 H1 H2 I2 J2 K1 K2 L1 L2 M2 N1 N2 O1 O2 P2 Q1	26
28	Q2	47	B2 D1 E1 E2 F2 G1 G2 H2 I1 I2 J1 J2 K2 L2 M2 N2 O1 O2 P1 P2 Q1 S1 S2	23
29	Q1	71	A1 C1 C2 D2 E1 E2 F2 G1 H1 H2 I1 I2 J1 J2 K1 L1 L2 M1 M2 N2 O1 O2 P1 P2	24
29	Q2	64	A1 C1 C2 D2 E1 E2 F1 F2 G1 H1 H2 I1 I2 J1 J2 K1 L1 L2 M1 M2 N2 O1 O2 P1 P2	25
31	Q1	87	A1 A2 B1 B2 C1 C2 D1 E1 F1 F2 G1 G2 H2 I2 J1 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1	29
31	Q2	75	A1 A2 B1 B2 C1 C2 D1 E1 F1 F2 G1 G2 H1 H2 I2 J1 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2	31
33	Q1	36	A1 A2 C1 D1 D2 E1 F1 F2 G1 G2 H2 J2 K1 K2 L1 L2 M1 M2 N1 N2 P1 P2 Q2 S1 S2	25
35	Q1	66	A2 C2 D1 D2 E1 E2 F2 G1 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 M2 N1 N2 O1 O2 P2 S2	25
35	Q2	79	A2 D1 D2 H1 I2 J2 K2 L1 N2 O2	10
36	Q1	47	E1 F1 G2 H1 H2 K1 L1 O2 P1	9
36	Q2	25	A1 A2 B1 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 S1 S2	33
37	Q2	38	A2 B2 C1 C2 D2 F1 F2 G1 G2 H2 I1 I2 J1 J2 K1 K2 L1 L2 N1 N2 O1 O2 P1 P2 S1 S2	26
38	Q1	63	C1 C2 D1 D2 E1 E2 F2 G1 I1 I2 N2	11
38	Q2	55	A2 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 I1 I2 J1 J2 K1 L1 L2 M1 M2 N2 O1 O2 P1 P2	26
39	Q2	82	B1 B2 C1 D1 D2 E1 E2 F1 F2 G1 G2 H2 I1 I2 J2 K1 K2 L1 L2 M1 M2 N2 O1 P1 P2 S1 S2	27
43	Q1	67	A1 A2 C2 D1 D2 E1 F2 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 L2 M2 O1 O2 P2	22
43	Q2	62	A1 A2 C2 D1 D2 E1 E2 F2 G2 H1 H2 I1 I2 J1 J2 K1 K2 L1 L2 M2 O1 O2 P2	23
44	Q2	35	A1 A2 B2 C1 C2 D1 E1 F2 G2 I1 J2 K1 K2 L1 L2 N2 P2 Q1	18
45	Q1	52	A2 D1 D2 E1 E2 I2 K1 K2 L1 L2 P2 S1 S2	13
45	Q2	45	A1 A2 B2 C1 C2 D1 D2 E1 E2 F2 H1 H2 I1 I2 K1 K2 L1 L2 N1 N2 O2 P2 S1 S2	24
46	Q2	44	A1 A2 C1 C2 D1 D2 E2 F1 F2 G1 G2 H1 H2 I1 I2 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 Q1 S1 S2	31
47	Q2	85	A1 A2 B1 B2 C1 D1 D2 F1 F2 G2 H1 I1 I2 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P2 S1 S2	27
49	Q2	41	A1 A2 C1 C2 D1 D2 E1 E2 G1 G2 H1 H2 I2 J2 K1 K2 L1 L2 M1 M2 N1 N2 O1 O2 P1 P2 Q1 S1 S2	29
50	Q2	80	A1 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 H2 I1 I2 J1 J2 K1 K2 M1 M2 O2 S1 S2	24

Table J.17: Cases where at least 1 review obtained higher agreements than expert A.

Summary. The majority of other reviews obtained significantly higher agreement than expert Q in 28 cases (36 reviews).

case	review	θ	Reviews which have significantly higher agreement	total
8	S1	65	A1 A2 B1 D2 E2 F2 G1 H2 I1 J2 K2 L2 M1 M2 N1 O1 P1 P2 Q2	19
8	S2	65	A1 A2 B1 D2 E2 F2 G1 H2 I1 J2 K2 L2 M1 M2 N1 O1 P1 P2 Q2	19
11	S1	65	A2 B2 C1 C2 D1 E1 E2 F2 G1 H1 H2 I1 I2 K1 L1 M2 N2 P2	18
11	S2	65	A2 B2 C1 C2 D1 E1 E2 F2 G1 H1 H2 I1 I2 K1 L1 M2 N2 P2	18
29	S1	79	A1 C1 C2 D2 E1 E2 H2 I1 I2 J2 L2 N2 O1 O2 P1	15
29	S2	79	A1 C1 C2 D2 E1 E2 H2 I1 I2 J2 K1 L1 L2 N2 O1 O2 P1	17
35	S1	68	A2 D1 D2 E1 E2 G1 H1 I2 J1 J2 K2 L1 N2 O2	14
38	S1	32	A2 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 I1 I2 J1 J2 K1 L1 L2 M1 M2 N2 O1 O2 P1 P2	26
38	S2	32	A2 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 I1 I2 J1 J2 K1 L1 L2 M1 M2 N2 O1 O2 P1 P2	26
44	S1	33	A1 A2 B2 C1 C2 D1 D2 E1 E2 F2 G1 G2 I1 I2 J2 K1 K2 L1 L2 M2 N1 N2 O1 O2 P2 Q1	26
44	S2	33	A1 A2 B2 C1 C2 D1 D2 E1 E2 F2 G1 G2 I1 I2 J2 K1 K2 L1 L2 M2 N1 N2 O1 O2 P2 Q1	26
48	S1	50	A2 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H1 J2 K1 K2 L1 N1 N2 P1 P2	22
48	S2	50	A1 A2 B1 B2 C1 C2 D1 D2 E1 E2 F1 F2 G1 G2 H1 J2 K1 K2 L1 N1 N2 P1 P2	23

Table J.18: Cases where at least 1 review obtained higher agreements than the system.

Summary. The majority of other reviews obtained significantly higher agreement than the System in 5 cases (10 reviews).

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S Max. = 16	Num. reviews recommending C/S. Max. = 32	Agreement with C/S specified in A1			δ_t System (segs)	Agreement with C/S specified in A2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
3	YES	1	14	24	11	4	3	2	11	3	4	1
4	NO	-	2	3	0	0	3	-	-	-	-	-
5	NO	-	6	9	3	0	2	-	-	-	-	-
6	YES	0	16	31	31	0	0	0	31	0	0	0
9	YES	2	16	32	24	8	0	1	5	14	13	1
10	YES	2	16	30	10	6	14	3	19	9	2	1
16	YES	4	16	31	2	2	10	-	4	4	0	-
17	YES	0	16	32	17	0	5	1	17	0	5	1
19	YES	3	13	24	13	3	7	3	9	6	8	0
24	YES	1	15	28	15	10	1	-	22	2	2	-
29	YES	2	16	32	31	1	0	1	12	18	2	1
32	YES	4	13	24	5	3	9	-	12	6	4	-
34	NO	-	4	5	4	1	0	-	-	-	-	-
35	YES	2	16	32	6	15	11	4	30	2	0	2
37	NO	-	15	25	5	0	1	-	-	-	-	-
38	NO	-	16	30	-	-	-	-	24	1	5	-
40	YES	1	16	29	13	4	5	6	17	2	6	5
41	NO	-	2	3	2	1	0	-	-	-	-	-
43	YES	1	15	30	25	4	1	2	23	4	3	1
44	YES	0	16	31	13	13	4	-	13	13	4	-
48	YES	0	15	25	6	1	15	9	6	1	15	9

Detailed assessment of the cases recommended for CS by expert A.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S Max. = 16	Num. reviews recommending C/S. Max. = 32	Agreement with C/S specified in B1			δ_t System (segs)	Agreement with C/S specified in B2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
2	YES	2	11	16	1	1	2	-	1	3	1	-
3	YES	3	14	24	1	0	3	8	3	2	12	5
4	NO	-	2	3	-	-	-	-	2	0	1	-
5	YES	1	6	8	1	2	4	-	2	1	2	-
6	YES	0	16	31	31	0	0	0	31	0	0	0
9	YES	1	16	32	26	5	1	2	23	9	0	1
10	YES	0	16	30	18	4	7	0	18	4	7	0
12	YES	4	3	3	1	0	2	-	2	0	0	-
13	YES	9	8	13	4	2	3	-	1	0	1	-
15	NO	-	1	2	1	0	0	-	-	-	-	-
16	YES	3	16	31	1	1	0	-	1	1	0	-
17	YES	1	16	32	5	4	2	7	8	2	9	6
18	YES	1	2	2	0	0	1	-	0	0	0	-
19	YES	1	13	24	12	3	8	3	8	5	3	4
22	NO	-	1	2	1	0	0	-	-	-	-	-
24	YES	3	15	28	1	4	14	-	16	10	0	-
29	YES	1	16	32	19	11	2	2	30	2	0	1
32	YES	0	13	24	7	8	8	-	7	8	8	-
34	YES	0	4	4	3	1	0	-	3	1	0	-
35	YES	1	16	32	21	10	1	3	5	16	11	4
37	YES	15	15	24	5	0	1	-	17	1	0	-

Detailed assessment of the cases recommended for CS by expert B.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S Max. = 16	Num. reviews recommending C/S. Max. = 32	Agreement with C/S specified in B1			δ_t System (segs)	Agreement with C/S specified in B2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4							
38	YES	1	16	29	4	12	13	-	16	13	0	-
40	YES	3	16	29	3	3	4	1	7	6	16	2
41	NO	-	2	3	2	1	0	-	-	-	-	-
43	YES	3	15	30	17	7	5	0	12	17	1	3
44	YES	0	16	31	13	13	4	-	13	13	4	-
48	YES	0	15	25	14	1	8	5	14	1	8	5

Detailed assessment of the cases recommended for CS by expert B (continued).

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S Max. = 16	Num. reviews recommending C/S. Max. = 32	Agreement with C/S specified C1			δ_t System (segs)	Agreement with C/S specified C2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
2	NO	-	11	17	0	0	0	-	-	-	-	-
3	YES	2	14	24	12	3	3	2	8	4	4	0
6	YES	0	16	31	31	0	0	0	31	0	0	0
9	YES	2	16	32	24	8	0	1	12	14	6	3
10	YES	0	16	30	18	4	7	0	18	4	7	0
13	NO	-	8	14	4	4	3	-	-	-	-	-
16	YES	9	16	31	2	3	3	-	4	4	1	-
17	YES	1	16	32	14	5	8	2	17	0	5	1
19	YES	5	13	24	9	5	3	4	8	2	11	1
24	YES	1	15	28	17	5	3	-	22	2	2	-
28	NO	-	0	0	0	0	0	-	-	-	-	-
29	YES	0	16	32	30	2	0	1	30	2	0	1
32	YES	1	13	24	9	6	4	-	11	6	5	-
35	YES	1	16	32	21	10	1	3	29	3	0	2
37	NO	-	15	25	-	-	-	-	17	1	0	-
38	YES	0	16	29	24	5	0	-	24	5	0	-
40	NO	-	16	30	-	-	-	-	18	2	0	5
43	YES	2	15	30	12	16	2	3	24	3	3	1
44	YES	1	16	31	19	4	7	-	13	13	4	-
48	YES	1	15	25	14	1	8	5	14	1	8	6

Detailed assessment of the cases recommended for CS by expert C.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS										
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS								
			Num. other experts recommending C/S max. = 16	Num. reviews recommending C/S. Max. = 32	Agreement with C/S specified in D1			δ_t System (segs)	Agreement with C/S specified in D2			δ_t System (segs)	
					Num. other reviews with C/S within (segs)	± 1	± 2		± 4	Num. other reviews with C/S within (segs)	± 1		± 2
2	NO	-	11	17	3			1		0		-	
3	YES	3	14	24	5	5	9	4	12	3	3	1	1
6	YES	1	16	31	26	5	0	1	31	0	0	0	0
7	NO	-	6	8	4	1	2	-	-	-	-	-	-
9	YES	0	16	32	11	15	6	3	11	15	6	3	3
10	YES	3	16	30	13	6	9	1	15	11	4	2	2
13	YES	4	8	13	2	2	4	-	4	4	2	-	-
16	YES	10	16	31	6	0	3	-	6	1	1	-	-
17	YES	2	16	32	4	12	13	4	9	1	9	6	6
19	YES	4	13	24	8	2	10	1	13	3	7	3	3
24	YES	1	15	28	15	10	1	-	6	11	11	-	-
29	YES	1	16	32	19	11	2	2	30	2	0	1	1
32	YES	1	13	24	4	3	10	-	3	1	8	-	-
35	YES	0	16	32	29	3	0	2	29	3	0	2	2
37	YES	15	15	24	5	0	1	-	17	1	0	-	-
38	YES	0	16	29	24	5	0	-	24	5	0	-	-
40	YES	1	16	29	10	12	5	4	7	9	13	3	3
43	YES	0	15	30	23	4	3	1	23	4	3	1	1
44	YES	3	16	31	20	4	6	-	7	8	13	-	-
48	YES	1	15	25	14	1	8	5	12	2	1	4	4

Detailed assessment of the cases recommended for CS by expert D.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max. = 16	Num. reviews recommending C/S. Max. = 32	Agreement with C/S specified in E1			δ_t System (segs)	Agreement with C/S specified in E2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
2	NO	-	11	17	1	3	2	-	-	-	-	-
3	YES	3	14	24	5	5	9	4	12	3	3	1
5	NO	-	6	9	1	3	4	-	-	-	-	-
6	YES	0	16	31	31	0	0	0	31	0	0	0
7	NO	-	6	8	4	1	2	-	-	-	-	-
9	YES	0	16	32	11	15	6	3	11	15	6	3
10	YES	2	16	30	13	5	10	1	19	9	2	1
12	NO	-	3	4	-	-	-	-	2	0	2	-
13	YES	6	8	13	2	2	5	-	2	2	4	-
16	YES	11	16	31	4	4	1	-	6	1	1	-
17	YES	1	16	32	8	2	17	5	8	2	9	6
19	YES	4	13	24	8	2	10	1	13	3	7	3
24	YES	1	15	28	15	10	1	-	6	11	11	-
29	YES	0	16	32	30	2	0	1	30	2	0	1
32	YES	0	13	24	11	4	9	-	11	4	9	-
35	YES	0	16	32	29	3	0	2	29	3	0	2
37	YES	1	15	24	3	1	1	-	4	0	1	-
38	YES	0	16	29	24	5	0	-	24	5	0	-
40	YES	4	16	29	18	2	5	5	7	4	9	1
43	YES	1	15	30	23	4	3	1	25	4	1	2
44	YES	2	16	31	14	12	4	-	7	7	14	-
48	YES	0	15	25	14	1	8	5	14	1	8	5

Detailed assessment of the cases recommended for CS by expert E.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in F1			δ_t System (segs)	Agreement with C/S specified in F2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
2	YES	17	11	16	0	0	0	-	0	4	1	-
3	YES	31	14	24	3	2	13	5	1	1	0	36
6	YES	0	16	31	31	0	0	0	31	0	0	0
9	YES	0	16	32	26	5	1	2	26	5	1	2
10	NO	-	16	31	-	-	-	-	15	11	5	2
12	NO	-	3	4	2	0	2	-	-	-	-	-
13	YES	1	8	13	3	2	3	-	1	2	4	-
16	YES	14	16	31	0	2	5	-	4	4	1	-
17	YES	1	16	32	17	0	5	1	9	8	3	0
19	NO	-	13	25	-	-	-	-	7	1	7	2
24	YES	1	15	28	17	5	3	-	22	2	2	-
29	YES	1	16	32	30	2	0	0	30	2	0	1
35	YES	0	16	32	25	5	2	1	25	5	2	1
37	YES	0	15	24	16	1	0	-	16	1	0	-
38	YES	1	16	29	23	1	5	-	24	5	0	-
40	YES	1	16	29	17	2	6	5	13	4	5	6
43	YES	3	15	30	2	4	22	5	26	4	0	2
44	YES	3	16	31	10	5	9	-	14	13	3	-
48	YES	0	15	25	14	1	8	5	14	1	8	5

Detailed assessment of the cases recommended for CS by expert F.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (secs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in G1			δ_t System (secs)	Agreement with C/S specified in G2			δ_t System (secs)
					Num. other reviews with C/S within (secs)				Num. other reviews with C/S within (secs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
3	NO	-	14	25	-	-	-	-	2	6	7	1
6	YES	0	16	31	31	0	0	0	31	0	0	0
7	NO	-	6	8	4	1	2	-	-	-	-	-
9	YES	1	16	32	26	5	1	2	11	15	6	3
10	YES	2	16	30	15	10	5	2	19	3	7	0
16	NO	-	16	32	-	-	-	-	6	2	1	-
17	YES	6	16	32	1	0	0	5	18	0	5	1
19	YES	0	13	24	7	2	11	1	7	2	11	1
24	YES	1	15	28	22	2	2	-	17	5	3	-
29	YES	1	16	32	30	2	0	1	30	2	0	0
32	NO	-	13	25	-	-	-	-	6	5	8	-
34	NO	-	4	5	-	-	-	-	4	1	0	-
35	YES	1	16	32	29	3	0	2	21	10	1	3
37	NO	-	15	25	17	1	0	-	-	-	-	-
38	YES	0	16	29	23	1	5	-	23	1	5	-
40	YES	6	16	29	14	4	5	6	6	3	6	0
43	YES	1	15	30	16	7	6	0	23	4	3	1
44	YES	0	16	31	13	9	7	-	13	9	7	-
48	NO	-	15	26	15	1	8	5	-	-	-	-

Detailed assessment of the cases recommended for CS by expert G.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in H1			δ_t System (segs)	Agreement with C/S specified in H2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
2	YES	0	11	16	0	0	0	-	0	0	0	-
3	YES	5	14	24	1	0	4	8	8	8	3	3
4	YES	0	2	2	1	0	1	-	1	0	1	-
5	YES	6	6	8	2	1	1	-	3	0	2	-
6	YES	1	16	31	31	0	0	0	30	1	0	1
7	NO	-	6	8	2	2	4	-	-	-	-	-
9	YES	0	16	32	11	15	6	3	11	15	6	3
10	YES	1	16	30	12	6	10	1	18	4	7	0
13	YES	1	8	13	4	2	4	-	3	4	3	-
15	YES	11	1	1	0	0	0	-	1	0	0	-
16	YES	0	16	31	0	5	5	-	0	5	5	-
17	YES	5	16	32	0	2	8	9	4	13	13	4
19	YES	2	13	24	3	5	7	5	13	2	8	3
22	YES	11	1	1	0	0	0	-	1	0	0	-
24	YES	3	15	28	3	5	20	-	1	1	2	-
29	YES	1	16	32	30	2	0	1	30	2	0	0
32	YES	4	13	24	4	1	7	-	12	4	8	-
35	YES	1	16	32	29	3	0	2	25	5	2	1
37	YES	20	15	24	0	0	5	-	18	0	0	-
38	YES	0	16	29	4	12	13	-	4	12	13	-

Detailed assessment of the cases recommended for CS by expert H.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in H1			δ_t System (segs)	Agreement with C/S specified in H2			
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	δ_t System (segs)						
40	YES	1	16	29	5	3	6	0	6	4	10	1
41	YES	3	2	2	0	2	0	-	2	0	0	-
43	YES	0	15	30	23	4	3	1	23	4	3	1
44	YES	2	16	31	0	0	7	-	5	2	13	-
48	YES	0	15	25	6	1	15	9	6	1	15	9

Detailed assessment of the cases recommended for CS by expert H (continued).

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in I1			δ_t System (segs)	Agreement with C/S specified in I2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
3	YES	0	14	24	11	4	3	2	11	4	3	2
6	YES	0	16	31	31	0	0	0	31	0	0	0
9	YES	0	16	32	23	9	0	1	23	9	0	1
10	YES	2	16	30	19	3	7	0	15	10	5	2
16	YES	0	16	31	5	0	3	-	5	0	3	-
17	YES	1	16	32	14	5	8	2	17	0	5	1
19	YES	0	13	24	11	4	9	2	11	4	9	2
24	YES	0	15	28	17	5	3	-	17	5	3	-
29	YES	0	16	32	30	2	0	0	30	2	0	0
32	NO	-	13	25	-	-	-	-	8	8	8	-
35	YES	0	16	32	29	3	0	2	29	3	0	2
37	NO	-	15	25	3	15	0	-	-	-	-	-
38	YES	0	16	29	24	5	0	-	24	5	0	-
40	YES	1	16	29	6	6	17	2	7	9	13	3
43	YES	1	15	30	25	4	1	2	23	4	3	1
44	YES	2	16	31	14	12	4	-	14	8	7	-
48	YES	0	15	25	6	1	15	9	6	1	15	9

Detailed assessment of the cases recommended for CS by expert I.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in J1			δ_t System (segs)	Agreement with C/S specified in J2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4		± 1	± 2	± 4			
2	NO	-	11	17	-	-	-	-	1	3	2	-
3	YES	28	14	24	0	0	0	30	12	4	3	2
6	YES	0	16	31	30	1	0	1	30	1	0	1
7	NO	-	6	8	0	2	2	-	-	-	-	-
9	YES	0	16	32	19	5	8	0	19	5	8	0
10	YES	2	16	30	19	9	2	1	10	6	14	3
16	YES	6	16	31	6	1	1	-	0	0	4	-
17	YES	4	16	32	15	5	7	2	9	2	8	6
18	NO	-	2	3	-	-	-	-	0	0	1	-
19	YES	2	13	24	12	3	9	2	9	4	3	4
24	YES	1	15	28	15	10	1	-	22	2	2	-
29	YES	1	16	32	30	2	0	0	30	2	0	1
32	YES	1	13	24	11	4	9	-	11	6	5	-
34	NO	-	4	5	0	5	0	-	-	-	-	-
35	YES	0	16	32	29	3	0	2	29	3	0	2
37	YES	1	15	24	16	1	0	-	17	0	0	-
38	YES	0	16	29	23	1	5	-	23	1	5	-
40	YES	4	16	29	7	6	16	2	14	4	4	6
43	YES	0	15	30	23	4	3	1	23	4	3	1
44	YES	3	16	31	7	8	13	-	20	4	6	-
48	NO	-	15	26	-	-	-	-	7	1	15	9

Detailed assessment of the cases recommended for CS by expert J.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in K1			δ_t System (segs)	Agreement with C/S specified in K2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
2	NO	-	11	17	-	-	-	-	3	1	0	-
3	YES	37	14	24	0	2	0	38	12	3	4	1
5	NO	-	6	9	3	1	2	-	-	-	-	-
6	YES	0	16	31	31	0	0	0	31	0	0	0
9	YES	1	16	32	19	5	8	0	23	9	0	1
10	YES	2	16	30	19	9	2	1	10	6	4	3
16	YES	2	16	31	5	1	1	-	4	2	1	-
17	YES	4	16	32	15	5	7	2	9	2	8	6
19	YES	2	13	24	8	1	11	1	7	12	5	1
24	YES	0	15	28	22	2	2	-	22	2	2	-
29	YES	2	16	32	31	1	0	1	12	18	2	1
32	YES	2	13	24	6	4	8	-	12	5	5	-
35	YES	1	16	32	21	10	1	3	29	3	0	2
37	YES	0	15	24	16	1	0	-	16	1	0	-
38	NO	-	16	30	25	5	0	-	-	-	-	-
40	YES	2	16	29	11	11	5	4	14	3	5	6
43	YES	1	15	30	23	4	3	1	25	4	1	2
44	YES	2	16	31	20	3	7	-	14	5	11	-
48	NO	-	15	26	-	-	-	-	2	21	1	7

Detailed assessment of the cases recommended for CS by expert K.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in L1			Agreement with C/S specified in L2				
					Num. other reviews with C/S within (segs)			δ_t System (segs)	Num. other reviews with C/S within (segs)			δ_t System (segs)
± 1	± 2	± 4	± 1	± 2	± 4							
3	NO	-	14	25	12	3	4	1	-	-	-	-
6	YES	0	16	31	31	0	0	0	31	0	0	0
7	YES	1	6	7	4	2	1	-	2	4	1	-
9	YES	0	16	32	23	9	0	1	23	9	0	1
10	YES	1	16	30	18	4	7	0	18	10	2	1
13	NO	-	8	14	-	-	-	-	4	4	3	-
16	YES	3	16	31	6	0	2	-	2	2	6	-
17	YES	0	16	32	14	5	8	2	14	5	8	2
19	YES	3	13	24	8	2	10	1	12	4	8	2
24	YES	1	15	28	17	5	3	-	22	2	2	-
29	YES	0	16	32	30	2	0	1	30	2	0	1
32	YES	2	13	24	6	6	8	-	12	3	9	-
35	YES	1	16	32	29	3	0	2	25	5	2	1
37	YES	0	15	24	16	1	0	-	16	1	0	-
38	YES	0	16	29	23	1	5	-	23	1	5	-
40	YES	1	16	29	17	2	6	5	10	12	5	4
43	YES	1	15	30	25	4	1	2	23	4	3	1
44	YES	2	16	31	14	5	11	-	20	3	7	-
48	YES	28	15	25	15	1	8	5	1	0	0	23

Detailed assessment of the cases recommended for CS by expert L.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in M1			δ_t System (segs)	Agreement with C/S specified in M2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4							
6	NO	-	16	32	31	1	0	1	-	-	-	-
9	YES	1	16	32	19	5	8	0	23	9	0	1
10	NO	-	16	31	-	-	-	-	1	6	10	5
16	YES	0	16	31	5	0	3	-	5	0	3	-
17	YES	4	16	32	1	0	2	4	10	8	2	0
19	NO	-	13	25	-	-	-	-	12	4	9	2
24	NO	-	15	29	-	-	-	-	18	5	3	-
29	YES	1	16	32	30	2	0	0	30	2	0	1
35	YES	1	16	32	9	16	7	0	25	5	2	1
38	YES	0	16	29	23	1	5	-	23	1	5	-
40	YES	0	16	29	13	4	5	6	13	4	5	6
43	YES	3	15	30	6	7	17	4	24	4	2	1
44	YES	0	16	31	13	9	7	-	13	9	7	-

Detailed assessment of the cases recommended for CS by expert M.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in N1			δ_t System (segs)	Agreement with C/S specified in N2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
2	YES	0	11	16	2	1	0	-	2	1	0	-
6	YES	0	16	31	31	0	0	0	31	0	0	0
9	YES	1	16	32	19	5	8	0	23	9	0	1
10	YES	1	16	30	6	4	16	4	9	7	14	3
13	NO	-	8	14	-	-	-	-	5	2	4	-
16	YES	2	16	31	4	2	1	-	5	1	1	-
17	YES	0	16	32	14	5	8	2	14	5	8	2
24	YES	1	15	28	17	5	3	-	22	2	2	-
29	YES	0	16	32	30	2	0	1	30	2	0	1
32	YES	2	13	24	3	2	10	-	10	5	4	-
35	YES	1	16	32	25	5	2	1	29	3	0	2
37	YES	1	15	24	17	0	0	-	16	1	0	-
38	YES	2	16	29	5	11	13	-	25	4	0	-
40	NO	-	16	30	14	4	5	6	-	-	-	-
43	YES	1	15	30	11	17	2	3	5	7	18	4
44	YES	1	16	31	9	5	10	-	13	9	7	-
48	YES	0	15	25	14	1	8	5	14	1	8	5

Detailed assessment of the cases recommended for CS by expert N.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in O1			δ_t System (segs)	Agreement with C/S specified in O2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
2	NO	-	11	17	1	4	1	-	-	-	-	-
3	YES	3	14	24	8	7	3	3	12	2	4	1
5	NO	-	6	9	3	1	2	-	-	-	-	-
6	YES	1	16	31	31	0	0	0	30	1	0	1
7	YES	3	6	7	2	2	3	-	4	1	1	-
9	YES	0	16	32	23	9	0	1	23	9	0	1
10	YES	1	16	30	18	4	7	0	12	6	10	1
16	YES	2	16	31	2	1	7	-	6	1	1	-
17	YES	4	16	32	9	2	16	5	18	0	4	1
24	YES	7	15	28	1	1	3	-	18	5	3	-
29	YES	1	16	32	30	2	0	1	30	2	0	0
32	YES	4	13	24	5	3	9	-	12	6	4	-
34	NO	-	4	5	4	1	0	-	-	-	-	-
35	YES	1	16	32	21	10	1	3	29	3	0	2
37	NO	-	15	25	17	1	0	-	-	-	-	-
38	YES	1	16	29	24	5	0	-	23	1	5	-
40	YES	5	16	29	3	3	5	1	11	12	5	4
43	YES	0	15	30	23	4	3	1	23	4	3	1
44	YES	0	16	31	6	8	14	-	6	8	14	-
48	YES	32	15	25	1	0	0	23	7	1	15	9

Detailed assessment of the cases recommended for CS by expert O.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in P1			δ_t System (segs)	Agreement with C/S specified in P2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
2	YES	2	11	16	0	0	0	-	0	0	1	-
3	NO	-	14	25	8	5	4	0	-	-	-	-
6	YES	0	16	31	31	0	0	0	31	0	0	0
9	YES	0	16	32	23	9	0	1	23	9	0	1
10	YES	2	16	30	19	9	2	1	10	6	14	3
16	YES	0	16	31	1	4	3	-	1	4	3	-
17	YES	1	16	32	9	8	3	0	17	0	5	1
19	YES	2	13	24	7	12	5	1	13	2	8	3
24	NO	-	15	29	-	-	-	-	23	2	2	-
29	YES	0	16	32	30	2	0	0	30	2	0	0
32	YES	1	13	24	5	5	8	-	9	6	4	-
35	YES	0	16	32	25	5	2	1	25	5	2	1
37	YES	0	15	24	16	1	0	-	16	1	0	-
38	YES	0	16	29	23	1	5	-	23	1	5	-
40	YES	1	16	29	17	2	6	5	13	4	5	6
43	YES	1	15	30	11	17	2	3	25	4	1	2
44	YES	2	16	31	2	7	13	-	14	8	7	-
48	NO	-	15	26	15	1	8	6	-	-	-	-

Detailed assessment of the cases recommended for CS by expert P.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 16	Num. reviews recommending C/S. Max = 32	Agreement with C/S specified in Q1			δ_t System (segs)	Agreement with C/S specified in Q2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
2	YES	6	11	16	1	2	2	-	0	2	2	-
3	NO	-	14	25	1	1	0	36	-	-	-	-
5	YES	12	6	8	0	0	0	-	2	1	1	-
6	YES	0	16	31	31	0	0	0	31	0	0	0
9	YES	1	16	32	26	5	1	2	23	9	0	1
10	YES	3	16	30	19	4	6	0	10	7	13	3
12	NO	-	3	4	1	0	2	-	-	-	-	-
13	YES	3	8	13	0	0	3	-	1	0	0	-
14	NO	-	0	0	0	0	0	-	-	-	-	-
16	YES	1	16	31	1	1	0	-	1	1	0	-
17	YES	2	16	32	6	3	2	7	9	1	17	5
18	NO	-	2	3	0	0	2	-	-	-	-	-
23	YES	8	0	0	0	0	0	-	0	0	0	-
29	YES	1	16	32	30	2	0	1	30	2	0	0
35	YES	0	16	32	25	5	2	1	25	5	2	1
37	NO	-	15	25	-	-	-	-	4	1	0	-
38	NO	-	16	30	-	-	-	-	5	12	13	-
40	NO	-	16	30	-	-	-	-	7	4	10	1
44	NO	-	16	32	14	9	7	-	-	-	-	-
48	NO	-	15	26	-	-	-	-	15	1	8	5

Detailed assessment of the cases recommended for CS by expert Q.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS							
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS					
			Num. other experts recommending C/S max = 17	Num. reviews recommending C/S. Max = 34	Agreement for C/S of S1			Agreement for C/S of S2		
					Num. other reviews with C/S within (segs)			Num. other reviews with C/S within (segs)		
± 1	± 2	± 4	± 1	± 2	± 4					
3	YES	0	15	26	9	5	4	9	5	4
6	YES	0	17	33	33	0	0	33	0	0
9	YES	0	17	34	21	5	8	21	5	8
10	YES	0	17	32	20	4	7	20	4	7
17	YES	2	17	34	11	8	3	16	5	8
19	YES	0	14	26	10	6	9	10	6	9
29	YES	0	17	34	32	2	0	32	2	0
35	YES	1	17	34	11	16	7	27	5	2
40	YES	0	17	31	7	3	6	7	3	6
43	YES	0	16	32	18	7	6	18	7	6
48	YES	0	16	27	0	0	1	0	0	1

Detailed assessment of the cases recommended for CS by the system.

Case Number	INTRA - AGREEMENT		INTER - AGREEMENTS									
	Was a C/S specified in both reviews ?	δ_t Intra - agreement (segs)	OVERALL - AGREEMENTS		TIMING OF AGREEMENTS							
			Num. other experts recommending C/S max = 17	Num. reviews recommending C/S. Max = 34	Agreement with C/S specified in R1			δ_t System (segs)	Agreement with C/S specified in R2			δ_t System (segs)
					Num. other reviews with C/S within (segs)				Num. other reviews with C/S within (segs)			
± 1	± 2	± 4	± 1	± 2	± 4	± 1	± 2	± 4				
2	YES	24	12	18	1	0	1	-	2	0	3	-
3	NO	-	15	26	0	0	0	15	-	-	-	-
4	NO	-	3	4	1	0	3	-	-	-	-	-
7	NO	-	7	9	0	0	0	-	-	-	-	-
8	NO	-	0	0	0	0	0	-	-	-	-	-
11	NO	-	0	0	0	0	0	-	-	-	-	-
13	NO	--	9	15	2	0	1	-	-	-	-	-
16	NO	-	17	33	1	1	2	-	-	-	-	-
21	NO	-	0	0	0	0	0	-	-	-	-	-
23	NO	-	1	2	-	-	-	-	1	0	0	-
31	NO	-	0	0	-	-	-	-	0	0	0	-
32	NO	-	14	26	-	-	-	-	0	0	0	-
34	NO	-	5	6	-	-	-	-	0	0	0	-
36	NO	-	0	0	0	0	0	-	-	-	-	-
39	NO	-	0	0	0	0	0	-	-	-	-	-
43	NO	-	16	32	2	16	11	1	-	-	-	-
44	NO	-	17	33	-	-	-	-	15	6	11	-
45	NO	-	0	0	-	-	-	-	0	0	0	-
49	NO	-	0	0	-	-	-	-	0	0	0	-

Detailed assessment of the cases recommended for CS by the plausible random numbers.