DEVELOPMENT AND EVALUATION OF A COMBINATION OF COMPUTER ANALYSIS OF CARDIOTOCOGRAPHY AND ELECTROCARDIOGRAPHY FOR INTRAPARTUM FETAL MONITORING

DESENVOLVIMENTO E VALIDAÇÃO DE UM SISTEMA DE ANÁLISE INTEGRADA DE CARDIOTOCOGRAFIA E ELECTROCARDIOGRAFIA FETAL INTRAPARTO

deceleration with reduced variability / Signal loss
DEVELOPMENT AND EVALUATION OF A COMBINATION OF COMPUTER ANALYSIS OF CARDIOTOCOGRAPHY AND ELECTROCARDIOGRAPHY FOR INTRAPARTUM FETAL MONITORING

DESENVOLVIMENTO E VALIDAÇÃO DE UM SISTEMA DE ANÁLISE INTEGRADA DE CARDIOTOCOGRAFIA E ELECTROCARDIOGRAFIA FETAL INTRAPARTO

PhD THESIS

DISSERTAÇÃO DE DOUTORAMENTO EM MEDICINA
Dissertação de candidatura ao grau de Doutor em Medicina, na área de Obstetrícia, submetida à Faculdade de Medicina da Universidade do Porto.

Orientador: Prof. Doutor Diogo Ayres-de-Campos

Co-orientador: Prof. Doutor Karl Rosén
To my sister Kiki
To my family and patients
CONTENTS

Abbreviations .................................................................................................................. 9

Summary .......................................................................................................................... 14
English language ............................................................................................................. 15
Portuguese language ....................................................................................................... 17

Chapter 1
Background ....................................................................................................................... 20
  Abstract ......................................................................................................................... 21
  Intrapartum fetal hypoxia ............................................................................................. 22
  Fetal acid-base balance ............................................................................................... 23
  Cardiotocography ......................................................................................................... 26
  Continuous ECG ST waveform analysis ..................................................................... 34
  Computerized CTG analysis ....................................................................................... 40

Chapter 2
Aim and outline of the thesis ......................................................................................... 52
  Brief overview ............................................................................................................. 53
  Aim and outline of the thesis ....................................................................................... 54

Chapter 3
Omniview-SisPorto® 3.5 - a central fetal monitoring station with on-line alerts based on
computerised cardiotocogram+ST events signals......................................................... 58

Chapter 4
Comparision of the evaluation of cardiotocographic events by Omniview-SisPorto® 3.5
and a consensus of clinicians ...................................................................................... 66

Chapter 5
Access to computerised analysis of intrapartum cardiotocographs improves clinician’s
prediction of newborn umbilical pH ......................................................................... 78

Chapter 6
Prediction of neonatal acidemia by computer analysis of fetal heart rate and ST event
signals ......................................................................................................................... 90

Chapter 7
A randomised clinical trial of intrapartum fetal monitoring with computer analysis
and alerts versus previously available monitoring – Study protocol ......................... 102

Chapter 8
General Discussion / Main Conclusions .................................................................. 116

List of publications ........................................................................................................ 124

Acknowledgments / Agradecimentos ........................................................................ 128
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAP</td>
<td>American Academy of Pediatricians</td>
</tr>
<tr>
<td>Accl</td>
<td>Acceleration</td>
</tr>
<tr>
<td>ACOG</td>
<td>American College of Obstetricians and Gynecologists</td>
</tr>
<tr>
<td>aLTV</td>
<td>Abnormal longterm variability</td>
</tr>
<tr>
<td>Ante</td>
<td>Antepartum</td>
</tr>
<tr>
<td>aSTV</td>
<td>Abnormal shortterm variability</td>
</tr>
<tr>
<td>BD</td>
<td>Base deficit</td>
</tr>
<tr>
<td>BD_{eff}</td>
<td>Base deficit in the extracellular fluid</td>
</tr>
<tr>
<td>BE</td>
<td>Base excess</td>
</tr>
<tr>
<td>BP</td>
<td>Biphasic event</td>
</tr>
<tr>
<td>BW</td>
<td>Birthweight</td>
</tr>
<tr>
<td>CTG</td>
<td>Cardiotocography</td>
</tr>
<tr>
<td>cCTG</td>
<td>Computerised cardiotocography</td>
</tr>
<tr>
<td>95%CI</td>
<td>95% confidence intervals</td>
</tr>
<tr>
<td>CP</td>
<td>Cerebral palsy</td>
</tr>
<tr>
<td>CT</td>
<td>Computer tomography</td>
</tr>
<tr>
<td>CTG</td>
<td>Cardiotocography</td>
</tr>
<tr>
<td>CS</td>
<td>Cesarean section</td>
</tr>
<tr>
<td>Decel</td>
<td>Deceleration</td>
</tr>
<tr>
<td>EFM</td>
<td>Electronic fetal monitoring</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal blood sampling</td>
</tr>
<tr>
<td>fECG</td>
<td>Fetal electrocardiogram</td>
</tr>
<tr>
<td>FHR</td>
<td>Fetal heart rate</td>
</tr>
<tr>
<td>HI</td>
<td>Hypoxic-ischaemic</td>
</tr>
<tr>
<td>HIE</td>
<td>Hypoxic-ischaemic encephalopathy</td>
</tr>
<tr>
<td>IA</td>
<td>Intermittent auscultation</td>
</tr>
<tr>
<td>ICC</td>
<td>Intraclass correlation coefficient</td>
</tr>
<tr>
<td>Intra</td>
<td>Intrapartum</td>
</tr>
<tr>
<td>Inst</td>
<td>Instrumental vaginal delivery</td>
</tr>
<tr>
<td>K</td>
<td>Kappa coefficient</td>
</tr>
<tr>
<td>LTV</td>
<td>Long term variability</td>
</tr>
<tr>
<td>MR</td>
<td>Magnetic resonance</td>
</tr>
<tr>
<td>N</td>
<td>Number of cases</td>
</tr>
<tr>
<td>NE</td>
<td>Not estimated</td>
</tr>
<tr>
<td>Neurol assess</td>
<td>Neurological assessment</td>
</tr>
<tr>
<td>NN</td>
<td>Neonatal</td>
</tr>
<tr>
<td>NICU</td>
<td>Neonatal intensive care unit</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>NPV</td>
<td>Negative predictive value</td>
</tr>
<tr>
<td>NR</td>
<td>Not reported in the studies</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>%a</td>
<td>Percentage of agreement</td>
</tr>
<tr>
<td>Pa</td>
<td>Proportion of agreement</td>
</tr>
<tr>
<td>Pathol</td>
<td>Pathologic</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive predictive value</td>
</tr>
<tr>
<td>Prol D</td>
<td>Prolonged deceleration</td>
</tr>
<tr>
<td>Psa</td>
<td>Proportions of specific agreement</td>
</tr>
<tr>
<td>Rep D</td>
<td>Repetitive decelerations</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomised controlled trial</td>
</tr>
<tr>
<td>RR</td>
<td>Relative risk</td>
</tr>
<tr>
<td>sd</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>Sens</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>Spec</td>
<td>Specificity</td>
</tr>
<tr>
<td>STV</td>
<td>Short term variability</td>
</tr>
<tr>
<td>Sus</td>
<td>Suspicious</td>
</tr>
<tr>
<td>UA</td>
<td>Umbilical artery</td>
</tr>
<tr>
<td>UC</td>
<td>Uterine contractions</td>
</tr>
<tr>
<td>V</td>
<td>Variability</td>
</tr>
<tr>
<td>Vag</td>
<td>Normal vaginal delivery</td>
</tr>
<tr>
<td>VP D</td>
<td>Very prolonged deceleration</td>
</tr>
<tr>
<td>W</td>
<td>Weeks</td>
</tr>
<tr>
<td>wK</td>
<td>Weight kappa coefficient</td>
</tr>
</tbody>
</table>
“...For fetal monitoring to be effective, the test must be performed correctly; its results must then be interpreted satisfactorily and finally, this interpretation must provoke an appropriate response.”

Grant A, 1989
Summary

Resumo
Summary

The ability to make an accurate assessment of fetal well-being during labour is a great challenge. Despite its widespread use in industrialised countries, cardiotocography (CTG) has a limited sensitivity and a high false positive rate, which leads to unnecessary obstetrical interventions and a consequent increase in maternal and neonatal risks. Among the causes of its limited effectiveness is the poor inter- and intraobserver agreement on fetal heart rate (FHR) interpretation and the limited capacity to predict neonatal outcome.

Chapter 1 describes the main short- and long-term consequences of intrauterine oxygen deprivation, as well as the clinical measures used for its assessment. This chapter also summarises the historical and scientific evidence on intrapartum fetal monitoring, including that on fetal electrocardiography (fECG) and central monitoring stations with computerised cardiotocogram (cCTG) analysis. The addition of ST waveform analysis improves the sensitivity of intrapartum FHR in detection of fetal acidemia, and has been shown to decrease the rate of metabolic acidosis at birth, while simultaneously reducing operative deliveries for fetal distress. In spite of this, reliance on visual CTG interpretation remains a major concern for the technology and has led to other developments such as cCTG analysis.

The main objective of this thesis was the development and evaluation of a technology, which integrates computer analysis of CTG with fECG ST waveform analysis (Omniview-SisPorto®), aimed at improving the sensitivity and specificity of intrapartum fetal monitoring. The ultimate goal is to increase its effectiveness, thereby reducing the incidence of adverse neonatal outcomes without increasing the number of operative deliveries for fetal distress (Chapter 2).

Omniview-SisPorto® is the first central monitoring system to incorporate cCTG analysis and ST event features, providing real-time alerts for healthcare professionals. The program’s main characteristics and an overview of the system’s alerts are described in chapter 3.

Chapter 4 describes an agreement study, where an evaluation of the basic CTG features identified by Omniview-SisPorto® was compared with a consensus of three clinicians. A satisfactory agreement between the two was obtained.

Chapter 5 describes a prospective study that evaluated the impact of providing access to the results of cCTG analysis on clinicians’ reproducibility and accuracy in prediction of neonatal outcome. Access to cCTG improved clinician’s interobserver agreement and their accuracy in prediction of newborn umbilical artery (UA) pH.
A prospective observational study was conducted to assess the accuracy of real-time alerts in prediction of severe neonatal acidemia, and this is described in chapter 6. A high sensitivity and specificity were obtained in this analysis, and all severely acidemic fetuses were detected by the system’s red alerts.

The possible benefit of introducing cCTG analysis, with or without joint ST analysis, into routine clinical practice requires the conduction of an adequately sized randomised controlled trial (RCT) that will start in 2010. The study design is presented in chapter 7.

Centralised monitoring of fetal signals with computer analysis of CTG and ST event features is a novel development in intrapartum assessment of fetal well-being, and several refinements are likely to be needed in the future. This thesis constitutes the initial evaluation of a promising technology developed with the aim of supporting clinical decision-making and reducing the complexity of fetal surveillance.
A avaliação do bem-estar fetal durante o trabalho de parto representa um dos maiores desafios da Obstetrícia. Apesar da sua extensa utilização em países industrializados, a cardiotocografia apresenta uma sensibilidade limitada e uma elevada taxa de falsos positivos, o que condiciona intervenções obstétricas desnecessárias com um consequente aumento dos riscos maternos e neonatais. Entre os motivos desta eficiência reduzida, contam-se uma elevada variabilidade inter- e intraobservador na interpretação dos traçados cardiotocográficos, bem como uma diminuída capacidade preditiva do desfecho neonatal.

O capítulo 1 descreve as principais consequências a curto e longo prazo da privação de oxigénio no meio intrauterino, bem como os parâmetros clínicos usados para a sua avaliação. Este capítulo descreve também sumariamente os aspectos históricos e a evidência científica existente sobre a monitorização fetal intraparto, incluindo a relativa à electrocardiografia fetal e às centrais de análise automatizada de cardiotocografia. A utilização da análise do segmento ST da electrocardiografia fetal melhora a sensibilidade da análise cardiotocográfica na detecção de acidemia fetal. A evidência científica revela que esta metodologia, diminui a incidência de acidose metabólica ao nascimento e simultaneamente reduz a taxa de partos instrumentados e cirúrgicos por estado fetal não-tranquilizador. Contudo, a fraca fiabilidade da análise visual da cardiotocografia permanece uma grande preocupação em relação a esta tecnologia, e levou ao desenvolvimento de outros métodos complementares para avaliação do bem-estar fetal, nomeadamente a análise automatizada de cardiotocografia.

O principal objectivo desta tese consistiu no desenvolvimento e avaliação dum sistema que integra a análise computorizada de cardiotocografia e do segmento ST da electrocardiografia fetal (Omniview-SisPorto®), visando melhorar a sensibilidade e especificidade da monitorização fetal intraparto. O seu propósito final é, assim, o de melhorar a eficiência da monitorização fetal intraparto, proporcionando assim uma redução da incidência de desfechos neonatais adversos, sem aumentar o número de partos instrumentados e cirúrgicos por estado fetal não-tranquilizador (capítulo 2).

A Omniview-SisPorto® é a primeira central de monitorização, que incorpora a análise computorizada de cardiotocografia e a análise do segmento ST, fornecendo alertas clínicos em tempo real. As características principais do sistema e uma descrição detalhada dos alertas são descritos no capítulo 3.

O capítulo 4 apresenta um estudo que compara a avaliação de parâmetros cardiotocográficos básicos realizado pela Omniview-SisPorto® com um consenso de clínicos, onde se obteve uma elevada concordância.
O capítulo 5 descreve um estudo prospectivo que avaliou o impacto da análise computorizada na reprodutibilidade e na acuidade dos clínicos relativa à previsão do desfecho neonatal. Os resultados revelam que o acesso à análise automatizada de cardiotocografia aumenta a concordância inter-observador e melhora a acuidade na previsão do pH do sangue da artéria umbilical.

No capítulo 6 descreve-se um estudo prospectivo observacional, cujo objectivo foi de avaliar a acuidade dos alertas do sistema na predição da acidemia neonatal grave. Obteve-se uma elevada sensibilidade e especificidade nesta avaliação, e todos os casos de acidemia neonatal grave foram detectados pelos alertas vermelhos do sistema.

O possível benefício da introdução da análise computorizada de cardiotocografia, com ou sem análise do segmento ST, na prática clínica diária requer um ensaio clínico randomizado com uma amostra populacional representativa. Este ensaio terá início em 2010 e o desenho do estudo encontra-se descrito no capítulo 7.

As centrais de monitorização fetal com análise computorizada de cardiotocografia e do segmento ST representam um novo desafio na avaliação do bem-estar fetal, que deve continuar a ser aperfeiçoado no futuro. A presente tese descreve o desenvolvimento e a avaliação clínica inicial desta tecnologia promissora, desenvolvida com o objectivo de sustentar a decisão clínica e de reduzir a complexidade da vigilância fetal.
Chapter 1

Background
ABSTRACT

Hypoxic-ischaemic (HI) brain injury, as a result of oxygen deficiency during term labour, represents a major issue in intrapartum care. In developed countries, it affects around 2.5 newborns in every 1000 live births and accounts for almost 15% of cases of cerebral palsy (CP). The degree of acidemia at birth is the most useful marker of fetal intrapartum hypoxia. It is routinely available, highly specific and has a reasonable prevalence in the population; therefore it serves as a biological marker and also as a potentially useful variable for quality control of obstetric management.

Continuous fetal monitoring is currently the most commonly used obstetric procedure around the world. The global impact of continuous CTG, assessed by studies of effectiveness, reveals a significant reduction in the incidence of neonatal seizures, but an increase in the rate of operative deliveries. CTG can be regarded, in the intrapartum setting, as a useful screening test for fetal hypoxia with a high negative predictive value (NPV), but with an unreasonably low positive predictive value (PPV). There is strong evidence of a marked intra- and interobserver variation in FHR interpretation, an aspect that may have important impact on the limitations of the technology.

The STAN methodology® is based on an integrated interpretation of visual analysis of the CTG and automated analysis of ST waveforms of the fECG. It is associated with a decreased incidence of severe metabolic acidosis at birth and fewer operative deliveries for fetal distress, when compared with isolated CTG. While there is evidence that the technology improves the sensitivity of intrapartum CTG, it retains an undesirably low PPV. Agreement on clinical management seems to be improved with this approach, but the reproducibility of FHR pattern classification remains moderate. This has led to the generalised view that visual CTG assessment remains the technology’s weakest aspect.

Computer systems of FHR interpretation have been developed to overcome this limitation of intrapartum fetal monitoring.
INTRAPARTUM FETAL HYPOXIA

The ultimate purpose of intrapartum fetal monitoring is the evaluation of fetal well-being and, in doing so, to distinguish the potentially compromised fetus from the healthy well-oxygenated one. Adverse events in early life are known to have an impact on both child development and long-term adult health. Perinatal asphyxic brain injury, in particular, constitutes a major clinical hazard and represents an important issue in intrapartum care.

Fetal asphyxia corresponds to a drastic decrease in oxygen supply to the fetus, which leads to severe hypoxia, metabolic acidosis and subsequently to cellular damage. Apart from this pathophysiological definition, there was a need to establish diagnostic criteria for intrapartum asphyxia. The criteria, which led to a greater consensus in the scientific community, were those elaborated by the International Cerebral Palsy Task Force in 1999, revisited with minor revisions by the American College of Obstetricians and Gynecologists (ACOG) and the American Academy of Pediatricians (AAP) in 2003.1,2 The following four criteria are required:

- Profound metabolic acidosis (pH<7.00 and base deficit ≥12 mmol/l) on umbilical artery (UA);
- Early onset of severe or moderate neonatal encephalopathy in infants born at 34 or more weeks of gestation;
- Cerebral palsy of the spastic quadriplegic or dyskinetic type;
- Exclusion of other identifiable etiologies such as trauma, coagulation disorders, infectious conditions and genetic disorders.

Additional criteria suggesting an intrapartum timing, but nonspecific to asphyxia, include a sentinel hypoxic event during labour, severe electronic fetal monitoring abnormalities, Apgar score of 0 to 3 beyond five minutes, onset of multisystem involvement within 72 hours of birth and early imaging studies showing evidence of an acute nonfocal cerebral abnormality1,2.

A few published cases do not support the choice of some of these criteria. For example the cut-off value for pH below 7.0 is not universally agreed upon, and some investigators have reported cases of intrapartum HI injury occurring in neonates with a cord blood pH above 7.03-7. Apgar scores can be affected by multiple factors other than hypoxia, including maternal sedation or anesthesia, congenital malformations, resuscitation efforts and the presence of infection8,9. On the other hand, although term infants with 5-minute Apgar scores below 4 have a 386-fold increased risk for neonatal death and a 81-fold increased risk of CP, only 7% of survivors will develop CP10. Neonatal neuroimaging has been shown to provide useful information on the type and timing of brain injury11-13. Various modalities have been used to evaluate the brain in the presence of neonatal encephalopathy, including cranial sonography, computed
tomography and magnetic resonance imaging with spectroscopy. Although these have become increasingly important in the study of neonatal encephalopathy, the images must be interpreted with great caution, as the downstream consequences of insults of varying etiologies can produce similar images, thus concurring with the clinical signs of neurological depression but not necessarily being etiologically specific\textsuperscript{14}.

Despite several efforts to reach a clinical consensus and to standardise the diagnosis and classification of intrapartum HIE, several definitions have been used for this entity, which has contributed to variations in reported results. Regardless of the definition used, it is clear that the vast majority of CP cases are not associated with intrapartum HIE\textsuperscript{15-17}. One study, evaluating a population of 6000 children with CP from 13 geographically defined populations in Europe born between 1980 and 1990, reported an overall rate of 2.08 per 1000 live births\textsuperscript{18}. The precise prevalence of CP in the United States is uncertain, because consistent information is lacking on follow-up of an entire population, especially for term and late preterm infants, which comprise the majority of births. A population-based surveillance study using data from three regions in the United States estimated a prevalence of 3.6 cases per 1000 children at eight years of age, but the study did not distinguish between children with or without a history of prematurity\textsuperscript{19}. No data is available on the prevalence of CP in North America for term infants born since the mid-1980s\textsuperscript{20}. The estimated proportions of the various etiological factors responsible for CP depend on the criteria used to define this entity. Wu et al, based on neuroimaging findings, reported the following numbers: 22% due to perinatal ischemic stroke, 15% associated with congenital malformations, and 12% due to HIE\textsuperscript{20}. Clinical studies have revealed that 11-12% are associated with intrauterine exposure to inflammation, 6% with birth asphyxia, and 5% with complications of multiple births\textsuperscript{21,22}.

Despite the referred limitations, the consensus statement has been useful to audit CP causation and to exclude cases of primary intrapartum hypoxia\textsuperscript{23}. In developed countries the incidence of HIE at term is around 2.5 per 1000 live births, and the proportion of cases of CP associated with intrapartum asphyxia is about 15%\textsuperscript{24}, but these numbers may underdiagnose intrapartum hypoxic CP, due to the mentioned limitations in the diagnostic criteria.

**FETAL ACID-BASE BALANCE**

Umbilical cord blood gas and pH values should be obtained routinely in high-risk deliveries and whenever newborn depression occurs, because they assist with clinical management and may exclude the diagnosis of birth asphyxia in approximately 80% of depressed term newborns\textsuperscript{25}. The process of obtaining and analysing cord samples for acid base data analysis has been extensively reviewed by Westgate et al\textsuperscript{26}. Sampling
umbilical venous blood alone is not recommended, because arterial blood is more representative of the fetal condition and because arterial acidemia may occur with a normal venous pH. The sampling technique is simple and easily mastered by any health care provider in the delivery room. Pre-heparinised syringes ensure a consistent dose of heparin. In clinical practice sometimes only one sample is acquired. When this occurs, it should be treated as a venous sample. Even when no cord acid base data are collected, the newborn’s acid-base status may still be evaluated, provided that a blood gas sample is obtained within the first hour of life.

The degree of metabolic acidosis at birth is generally considered to be an objective and useful marker of fetal intrapartum hypoxia. It is easily obtainable and cases of severe acidemia have a reasonable prevalence in the general population; therefore it serves as a biological marker and also as a potentially useful variable for quality control of obstetric management.

The cut-off value of 7.0 for defining severe acidemia is not universally agreed upon, and some researchers have reported cases of HI injury in neonates with a cord pH above this value. An UApH value <7.05 identifies fetuses adjusting to hypoxia and is associated with neonatal complications, but a statistically significant increase in the incidence of serious neonatal morbidity is not seen until the UApH level is <7.00. The risk of complications, such as neurological damage, increases when tissue oxygen levels are sufficiently impaired to cause metabolic acidosis (indicating asphyxia), with the cut-off level of base deficit (BD) ≥ 12 mmol/l. Graham et al reviewed the scientific literature in order to examine the role of intrapartum hypoxic injury in causation of neonatal encephalopathy in non-anomalous term infants. The combined data of seven studies found that the incidence of UApH <7.0 is 3.7 (range: 2.9-8.3/1000) per 1000 term live births. Of the neonates with this degree of acidosis, 23.1% suffered neonatal neurological morbidity or mortality, 17.2% (range: 5.1-30.3%) survived with neurologic morbidity, 16.3% (range: 10.3-24.8%) developed seizures and 5.9% (range: 4.3-9.5%) died during the neonatal period. In Europe the incidence of metabolic acidosis ranges from 0.06 to 5%.

To optimally understand the fetal acid-base state, it is important to review the physiologic principles underlying fetal acid-base balance. Under normal physiologic conditions, the fetus produces both volatile (CO2) and nonvolatile acid. Volatile acids are usually transferred rapidly across the placenta by diffusion into the maternal blood. This process requires adequate blood flow on each side of the placenta. During labour, the nonvolatile acid, lactic acid is produced as a compensatory mechanism to hypoxia. Acid-base balance is maintained by initial buffering of the acid load and subsequent excretion in the urine. Buffers can be found in the extracellular and intracellular fluids and in bone. Intracellular buffers are proteins, organic and inorganic phosphates, and hemoglobin. Initial buffering is primarily achieved by bicarbonate (HCO3) in the
extracellular fluid and then by proteins and phosphate in the cells. The net effect of buffering by HCO$_3^-$ is a fall in serum HCO$_3^-$ concentration. Buffering capacity is the amount of acid that can be added or subtracted to cause a change in pH of one unit. Small changes in pH may have clinical significance, as the buffering capacity has been reduced. The concept of base excess (BE) was introduced as a measure of magnitude of the metabolic acid-base change (i.e., fixed acids) from normal, even in the presence of a concomitant respiratory acidemia (i.e., carbon dioxide pressure change). BE is calculated from the measured values of pH and carbon dioxide partial pressure, by means of an acid-base chart. BD is the inverse value of base excess.

Carbon dioxide is produced in large amounts in the cellular energy-yielding metabolic processes, and a continuous placental blood flow is required to avoid carbon dioxide accumulation. When gas exchange is acutely compromised, carbon dioxide concentration in fetal circulation increases and formation of carbonic acid and bicarbonate occurs. Some of the hydrogen ions in carbonic acid become free and lower the blood pH (respiratory acidemia). These acid-base changes are usually transitory, and resolve rapidly with re-establishment of placental gas exchange. However, if an important decrease in fetal oxygen concentration occurs, oxygen consumption will decrease, blood flow will be redistributed to spare the most vital organs and anaerobic metabolism will be initiated to maintain energy production, albeit with decreased efficiency. Under these conditions, lactate (an end product of anaerobic metabolism) is produced, resulting in metabolic acidosis. Unlike carbon dioxide, lactate is expelled slowly from the fetus. Respiratory and metabolic acidosis have different etiologies and therefore imply distinctive prognosis, although they can of course occur simultaneously. Metabolic acidosis carries with it a risk of cellular damage and reduced organ function, requires time to develop and persists for longer periods of time.

In order to assess the acid-base balance of the fetus, more information is needed than the UApH. Blood gas analysers usually provide measurements of pH, PCO$_2$ and bicarbonate concentrations on umbilical artery and vein blood samples. Empirical data and mathematical modeling of the buffer composition has provided us with models on how the buffers are behaving in the body. After initial work by Stow and Astrup, Siggaard-Andersen published in 1963 the blood acid-base alignment normogram. This required pH to be measured at known PCO$_2$ levels. These initial algorithms were found to overestimate the metabolic acidosis component in cases of mixed acidemia and an improved acid-base chart was introduced by Siggaard-Andersen in 1971, assessing the distribution of buffers in the whole extracellular fluid. Unfortunately, blood gas machines have not adopted the acid-base chart algorithms, with the implication that metabolic acidosis becomes overestimated, as a high PCO$_2$ causes a falsely low BE. Values obtained from these machines need to be recalculated in order to obtain a rigorous assessment of acid-base state.
The BD in the extracellular fluid (BD_{ecf}) in the umbilical artery and vein serves as a marker of the duration of hypoxia. A high BD_{ecf} in the cord artery combined with a normal value in the cord vein indicates a short lasting hypoxic process. On the contrary, when high levels are reached in both umbilical artery and vein, the underlying process of hypoxia has lasted long enough for an equilibrium to be reached in both vessels. A complete blood gas analysis thus provides important information on the type of acidemia (respiratory/mixed versus metabolic) and on the duration of the event (acute versus chronic).

**CARDIOTOCOGRAPHY**

**HISTORICAL ASPECTS**

The first description of fetal heart sounds dates from the 18th century, but it was only with Kergeradee in Europe, at the beginning of 19th century, that this discovery was seen as a method for assessing fetal well-being. Fetal auscultation with the aid of a stethoscope, developed over the course of 19th century, and Pinard’s version of the fetal stethoscope was first described in 1876. By the beginning of the 20th century auscultation of the fetal heart had become an established practice in Europe.

The first continuous register of FHR, using a microphone placed on the maternal abdomen, was accomplished in 1891. Subsequent improvements in this technique led to the introduction of phonocardiography in the late 1950s. External fetal electrocardiography was described for the first time by Cremer in 1906. Internal fetal electrocardiography, using a fetal scalp electrode was introduced by Hon in 1960 and was incorporated into routine clinical practice in 1974. The use of ultrasound Doppler for the evaluation of FHR was first described in 1964 and, because of its non-invasive nature and generally good quality of acquisition, was rapidly incorporated into commercial devices, thus replacing phonocardiography. Ultrasound Doppler and internal electrocardiography remain as the most widely used techniques for intrapartum continuous FHR evaluation.

In the 1950s Smith introduced the external tocodinamometer for assessment of uterine contractions, and an internal probe for measurement of intrauterine pressure was also developed in that decade.

Cardiotocography is said to have been developed by Caldeyro-Barcia, Hon and Hammacher in the late 1950s, as a technique for the simultaneous acquisition of FHR, uterine contractions and fetal movements. The first commercial device was introduced in 1968. The methodologies used for the acquisition and processing of the three
cardiotocographic parameters have changed over time, with a profound impact on the quality and interpretation of tracings.

External CTG monitoring currently extracts the FHR signal using ultrasound Doppler and processes it by autocorrelation. Internal FHR monitoring is a more invasive procedure limited to the intrapartum period, which uses a bipolar spiral electrode inserted transcervically into the fetal scalp and a reference electrode placed on the maternal thigh. The internal electrode detects the fECG and calculates FHR based on the interval between R waves. This signal provides accurate measurement of all FHR parameters, including beat-to-beat variability. Artifacts are rare, and autocorrelation is unnecessary.

Currently, continuous fetal monitoring is the most common obstetric procedure in the world. However, like intermittent auscultation, it was introduced into the clinical setting before its effectiveness had been fully evaluated.

**SCIENTIFIC VALIDATION**

**VALIDITY**

The validity of a method is defined as its capacity to detect or to foresee a pre-defined clinical condition and is quantified by the method’s sensitivity, specificity, positive and negative predictive values and likelihood ratios. An overview of the results of published studies evaluating the validity of CTG is presented in table I.
Table I: Validity of CTG by author and year of publication.  

<table>
<thead>
<tr>
<th>FHR Pattern</th>
<th>Terminology</th>
<th>Result</th>
<th>NN Outcome</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Presence of accelerations</strong></td>
<td>Spec: 97%</td>
<td>Apgar 5'&gt;7</td>
<td>Krebs, 82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NPV: 84.99%</td>
<td>Apgar 5'&gt;7 pH&gt;7.20</td>
<td>Samueloff, 94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PPV: 98%</td>
<td>Apgar 5'&gt;7 pH&gt;7.15</td>
<td>Parer, 2006</td>
<td></td>
</tr>
<tr>
<td><strong>FHR variability</strong></td>
<td>NPV: 92.7%</td>
<td>Apgar 1'&gt;5&gt;7</td>
<td>Berkus, 99</td>
<td></td>
</tr>
<tr>
<td><strong>Reassuring/Normal</strong></td>
<td>NPV: 96.7%</td>
<td>pH&gt;7.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Absence accelerations</strong></td>
<td>Sens: 92.3%; Spec: 61.7%</td>
<td>HIE</td>
<td>Larma, 2007</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PPV: 2.7%; NPV: 82.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Loss of FHR variability</strong></td>
<td>Sens: 18%</td>
<td>Apgar 5'&lt;7</td>
<td>Samueloff, 94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spec: 79.8%; PPV: 26.9%; NPV: 92.6%</td>
<td>HIE</td>
<td>Larma, 2007</td>
<td></td>
</tr>
<tr>
<td><strong>Prolonged bradycardia</strong></td>
<td>OR: 3.6 95CI [1.2;11]</td>
<td>Apgar 5'&lt;7</td>
<td>Berkus, 99</td>
<td></td>
</tr>
<tr>
<td><strong>Severe, variable decelerations</strong></td>
<td>OR: 2.4 95CI [1.2;4]</td>
<td>pH&lt;7.15</td>
<td>Berkus, 99</td>
<td></td>
</tr>
<tr>
<td><strong>Late decelerations</strong></td>
<td>OR: 6.9 95CI [2.1;23]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Severe, variable or late deceleration</strong></td>
<td>Sens: 94%; Spec: 56%</td>
<td>Neurol assess 12months</td>
<td>Painter, 78</td>
<td></td>
</tr>
<tr>
<td><strong>Prolonged deceleration</strong></td>
<td>Spec: 95%</td>
<td>Apgar 1'&lt;7</td>
<td>Özden, 99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spec: 96.3%</td>
<td>Apgar 5'&lt;7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spec: 97.5%</td>
<td>pH&lt;7.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Loss of variability during deceleration</strong></td>
<td>Sens: 66.7%</td>
<td>Apgar 1'&lt;7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spec: 72.3%</td>
<td>Apgar 5'&lt;7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spec: 63.9%</td>
<td>pH&lt;7.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Taquicardia</strong></td>
<td>PPV: 10.29%</td>
<td>pH&lt;7.20</td>
<td>Gilstrap, 84, 87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PPV: 30.39%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bradycardia</strong></td>
<td>Sens: 15.4%</td>
<td>HIE</td>
<td>Larma, 2007</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spec: 98.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PPV: 66.7%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NPV: 89.4%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sens: 13.8-34.7%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spec: 89.1-91.3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PPV: 0.05-0.25%</td>
<td>CP</td>
<td>Nelson, 96</td>
<td></td>
</tr>
<tr>
<td><strong>Nonreassuring/abnormal</strong></td>
<td>Sens: 17.93%; Spec: 29.89%</td>
<td>Acidosis&gt;16mmol/l</td>
<td>Low, 99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sens: 68%; Spec: 71%</td>
<td>Apgar 5'&lt;7</td>
<td>Dellinger, 2000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PPV: 5%; NPV: 99%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sens: 100%; Spec: 60%</td>
<td>pH&lt;7.0</td>
<td>Amer-Wahlén, 2005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PPV: 3%; NPV: 100%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PPV: 31.50%</td>
<td>pH&lt;7.0</td>
<td>Williams, 2002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spec: 80%</td>
<td>pH&lt;7.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PPV: 23%</td>
<td>BD&gt;12mmol/l</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sens: 7.7%; Spec: 98.9%</td>
<td>pH&gt;7.15</td>
<td>Parer, 2006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PPV: 50%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NPV: 88.6%</td>
<td>HIE</td>
<td>Larma, 2007</td>
<td></td>
</tr>
</tbody>
</table>

(Legend: Cat – category; Sens – sensitivity; Spec – specificity; NN – neonatal; FHR – fetal heart rate; HIE – hypoxic-ischaemic encephalopathy; NVP – negative predictive value; PPV – positive predictive value; OR - odds ratio; Neurol asses – neurological assessment until 12 months of age)
Methodological limitations can be pointed out in many of these studies. Some evaluated a very small sample size\textsuperscript{39,42,46,47,51} and consequently had insufficient data to calculate sensitivity and specificity. Others did not define the study population in detail\textsuperscript{39,46} or had unbalanced cohorts and/or selection bias\textsuperscript{40,41,43,48}. The selected study outcomes (or cut-off values used to define an outcome) varied among the different studies, making comparisons difficult. Perhaps the most relevant limitation of these studies is the lack of uniformity in FHR pattern analysis and CTG classification. Finally, the time lag between tracing-end and delivery varied among the different studies, raising the possibility that hypoxic situations may have occurred after monitoring had ceased. It cannot be expected that the CTG will predict the fetal state accurately after a long period, where monitoring was not carried out, especially in an unstable period such as the second stage of labour.

The sensitivity and specificity of a test, in association with the prevalence of the target condition, dictate its positive and negative predictive values. CTG appears to be a highly sensitive test for conditions that have a low prevalence, such as metabolic acidosis. On the other hand, it has a high false-positive rate and, hence, a poor PPV. Overreacting to FHR patterns can result in an increase in unnecessary obstetric interventions, namely instrumental or cesarean delivery. When aiming to increase the specificity of CTG, by only reacting to grossly abnormal CTG patterns, the test becomes falsely reassuring in an important number of cases, with a resulting reduction in its sensitivity, \textit{i.e.} a reduction in the detection of potentially compromised fetuses.

Because of its high specificity, when less restrictive abnormality criteria are used, CTG can be regarded as a useful screening test for fetal hypoxia, but this relies on the need to employ other methods to identify compromised fetuses more accurately, and simultaneously avoid unnecessary interventions.

**EFFECTIVENESS**

CTG was introduced into clinical practice in the 1960s with the purpose of reducing perinatal mortality and long-term neurological sequelae. However, it was only in the 1970s that clear ideas were established on how the effectiveness of any technology or treatment should be evaluated.

CTG has been compared with intermittent auscultation (IA) in a total of 12 randomised or quasi-randomised controlled trials, which have been evaluated in several meta-analyses or systematic reviews, the last of which by \textit{Alfirevic et al}\textsuperscript{53}. These studies were conducted between 1976 and 1993 and involved 37,615 participating women\textsuperscript{54-64}. Two of these trials were methodologically considered of high-quality: the Melbourne study, published in 1976 by \textit{Renou et al}, and the Dublin study published in 1985 by
McDonald et al. Participants were considered to have a low-risk pregnancy in three studies (Melbourne 1981, Dallas 1986 and Lund 1994) involving a total of 19,651 women. They were considered to have a high-risk pregnancy in five studies (Denver 1976, Denver 1979, Melbourne 1976, Seattle 1987 and Pakistan 1989), involving 2109 participants. In four studies, both low- and high-risk pregnancies were included (Sheffield 1978, Dublin 1985, Copenhagen 1985 and Lund 1994), with a total of 15,865 women. Of these 12 trials, only two (Melbourne 1976 and Athens 1993) conducted staff training in CTG interpretation, before the start of the trial.

The main results of this systematic review are considered in Table II.

Table II: Comparison of outcomes with the use of continuous CTG vs. IA (adapted from Ref 53).

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Overall</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perinatal death</td>
<td>0.85 [0.59;1.23]</td>
<td>1.02 [0.61;1.71]</td>
<td>1.02 [0.31;3.31]</td>
</tr>
<tr>
<td>Cerebral palsy</td>
<td>1.74 [0.97;3.11]</td>
<td>2.54 [1.10;5.86]</td>
<td>1.02 [0.31;3.31]</td>
</tr>
<tr>
<td>HI encephalopathy</td>
<td>0.46 [0.04;5.03]</td>
<td>NE</td>
<td>NE</td>
</tr>
<tr>
<td>Neonatal seizures</td>
<td>0.50 [0.31;0.80]</td>
<td>0.66 [0.36;1.22]</td>
<td>0.36 [0.16;0.81]</td>
</tr>
<tr>
<td>Cord acidosis (pH&lt;7.10)</td>
<td>0.92 [0.27;3.11]</td>
<td>NE</td>
<td>NE</td>
</tr>
<tr>
<td>5-minute Apgar &lt;7</td>
<td>0.97 [0.72;1.31]</td>
<td>0.75 [0.33;1.70]</td>
<td>1.37 [1.01;1.87]</td>
</tr>
<tr>
<td>Admission to NICU</td>
<td>1.01 [0.93;1.10]</td>
<td>2.02 [1.58;2.57]</td>
<td>1.95 [0.91;4.18]</td>
</tr>
<tr>
<td>CS rate</td>
<td>1.66 [1.30;2.13]</td>
<td>2.02 [1.58;2.57]</td>
<td>1.95 [0.91;4.18]</td>
</tr>
<tr>
<td>CS rate for abnormal FHR/acidosis</td>
<td>2.37 [1.88;3.00]</td>
<td>2.46 [1.69;3.59]</td>
<td>2.31 [1.49;3.59]</td>
</tr>
<tr>
<td>Instrumental birth</td>
<td>1.16 [1.01;1.32]</td>
<td>1.03 [0.85;1.26]</td>
<td>1.29 [1.02;1.62]</td>
</tr>
<tr>
<td>Instrumental birth for abnormal FHR/acidosis</td>
<td>2.54 [1.95;3.31]</td>
<td>NE</td>
<td>NE</td>
</tr>
</tbody>
</table>

(Legend: HI – hypoxic-ischaemic; NICU – neonatal intensive care unit; CS – cesarean section; FHR – fetal heart rate; NE – not estimated; Results in relative risks (RR); results with 95% interval confidence (95IC) [a,b]; RR >1 favours IA, RR<1 favours CTG)

No significant difference in perinatal mortality was found between the two arms. In a former meta-analysis carried out by Vintzileos et al., a subgroup analysis of perinatal mortality was undertaken, and a significant reduction in perinatal deaths due to hypoxia was found in the CTG arm [RR=0.41 (95CI 0.17-0.98)]. This was however, a post-hoc analysis that is prone to selection bias.

In three of the 12 RCTs a subsequent follow up study of the newborns was conducted, but no difference in the incidence of CP between the arms was documented. Data on the incidence of CP are heavily influenced by one trial (Seattle
1987) with 386 participants that only enrolled infants born at less than 32 weeks (W) of gestational age and found an increased incidence of CP in the CTG arm. Management after the detection of FHR abnormalities was not consistent in both arms of this trial, and a significantly longer mean delay between the onset of abnormal FHR patterns and birth was recorded in the CTG arm (105 minutes versus 45 minutes in the IA arm). The other two studies found no differences between the arms in the incidence of CP at the end of the follow-up period.

The use of continuous CTG monitoring was associated with a 50% reduction in the incidence of neonatal seizures. This finding was consistent across the trials and subgroups, although the incidence of neonatal seizures varied considerably between the trials. Although continuous CTG had a beneficial impact on this adverse neonatal outcome, a long-term follow-up of the children who survived these neonatal seizures in the Dublin trial, showed that they were not associated with CP69.

No differences were seen in the incidence of low Apgar scores or admissions to neonatal intensive care units (NICU), cord blood gas values or HIE. The latter two outcomes were evaluated in a small number of studies. Umbilical blood gases values were determined in eight trials (but not in all participants) and only five presented values for both umbilical artery and vein. In the other three, only blood from one vessel was obtained.

Operative delivery rates (instrumental vaginal delivery and cesarean section [CS] rates) were significantly increased in the CTG group, as compared with the IA. This effect was more pronounced when intervention was performed for presumed fetal distress. CS rates increased 66%, but there was substantial heterogeneity in this finding, and overall CS rates varied between 2.3% in the Dublin study (1985) and 35% in the Pakistan study (1989). A post-hoc analysis, comparing the effects on CS rates in trials with overall values lower or higher than 10%, uncovered a statistically significant test of interactions, suggesting that the adverse impact of CTG may be greater when baseline CS rates are high. There was also some evidence that trial quality influenced the size of the effects, i.e. the increase in CS rates appeared greater in studies of lower methodological quality.

From a numbers-need-to-treat perspective, it requires 628 women to be continuously monitored by CTG, to obtain one less neonatal seizure, but at the cost of an excess 11 CS..

It is important to consider that most of these trials were conducted at a time when many of the technological developments of CTG were still under way, there was very little consensus on FHR interpretation, and clinical experience with the technique was generally limited. In particular, there had been little recognition of the importance of FHR variability in CTG interpretation.
With such low prevalence of evaluated outcomes, namely perinatal mortality, CP and neonatal encephalopathy, any screening test would require a very high specificity in order to accomplish a significant reduction in their incidence. The trials are clearly underpowered to detect a difference in the incidence of perinatal mortality and CP.  

Finally, it is necessary to take into consideration that a diagnostic test will, by itself, not avoid an adverse outcome. Therefore, a RCT designed to evaluate a diagnostic test will mainly be evaluating the team’s capacity to interpret the test, the effectiveness of a pre-established management protocol (if existent), and the team’s capacity to manage the cases adequately.

LIMITATIONS OF CARDIOTOCOGRAPHY

DIFFERENCES IN CTG INTERPRETATION

The complexity of FHR patterns, particularly in the intrapartum, makes standardisation difficult. In the early years of CTG development, multiple classification systems for FHR interpretation were developed in different centres. Subsequently, some national and international organizations convened expert meetings to develop consensus guidelines on the terminology and classification of CTG, as well as to establish management recommendations. These definitions are primarily developed for visual interpretation of FHR patterns and can be applied to either internal or external FHR monitoring. The goal of these definitions and guidelines was to allow the predictive value of monitoring to be assessed more meaningfully and to allow evidence-based clinical management of intrapartum fetal compromise.

Despite these worthy efforts, important differences still exist on FHR nomenclature, tracing classification and management decisions based on CTG monitoring.

REPRODUCIBILITY

Visual analysis of tracings remains the most frequently used method of CTG evaluation in everyday clinical practice. Even in studies where standardisation of FHR interpretation criteria was achieved, important intra- and interobserver disagreement has been documented. Disagreements occur over the identification of individual CTG features, such as FHR baseline, accelerations, decelerations and contractions, as well as in overall classification of tracings and in clinical management based on CTG interpretation. Several motives for this phenomenon can be enumerated. Firstly,
ambiguous definitions of certain FHR parameters are still employed, such as for the baseline\textsuperscript{80-82}, and for classification of decelerations\textsuperscript{83-86}. Secondly, visual limitations are probably important in the evaluation of more detailed aspects, such as FHR variability. Lastly, limitations regarding the systematic and disciplined evaluation of FHR tracings are probably also of relevance.

Table III summarises the results of studies that evaluated intra- and interobserver variability on analysis of intrapartum CTG parameters. Different measures for quantifying agreement were used, namely the proportions of agreement (Pa), the kappa statistic (k), the weighted kappa statistic (wk) and the intraclass correlation coefficient (ICC). A k value exceeding 0.75 is considered to convey an excellent agreement, values between 0.40 and 0.75 a moderate to good agreement and values under 0.40 a poor agreement\textsuperscript{87,88}. If the 95% confidence interval for the proportions of agreement includes 0.5, and an adequate sample size is evaluated, then agreement should be considered poor\textsuperscript{88}. Values of the ICC exceeding 0.75 are generally interpreted as corresponding to an acceptable agreement\textsuperscript{89,90}. A detailed description of the advantages and limitations of these measures is beyond the scope of this work.

Table III. Studies evaluating CTG reproducibility (by first author, year of publication)\textsuperscript{81,84-86,91-102}.

<table>
<thead>
<tr>
<th>Identification of CTG features</th>
<th>Agreement</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Poor-good</td>
<td>Nielsen,87</td>
</tr>
<tr>
<td>Pa: 0.63-0.80</td>
<td></td>
<td>Lidegaard, 92</td>
</tr>
<tr>
<td>K: 0.16-0.56</td>
<td></td>
<td>Donker,93</td>
</tr>
<tr>
<td>ICC: 0.93</td>
<td></td>
<td>Todros,96</td>
</tr>
<tr>
<td>Accel/decel</td>
<td>Poor-moderate</td>
<td>Bernardes, 96,97</td>
</tr>
<tr>
<td>Pa: 0.27-0.60</td>
<td></td>
<td>A.-de-Campos,99</td>
</tr>
<tr>
<td>K: 0.03-0.64</td>
<td></td>
<td>Taylor,2000</td>
</tr>
<tr>
<td>ICC: 0.27-0.93</td>
<td></td>
<td>Devane, 2005</td>
</tr>
<tr>
<td>Variability</td>
<td>Poor-moderate</td>
<td>Nielsen,87</td>
</tr>
<tr>
<td>Pa: 0.13-0.64</td>
<td></td>
<td>Lidegaard, 92</td>
</tr>
<tr>
<td>K: 0.13-0.35</td>
<td></td>
<td>Paneth,93</td>
</tr>
<tr>
<td>Overall</td>
<td>Poor-good</td>
<td>A.-de-Campos,99</td>
</tr>
<tr>
<td>Pa: 0.18-0.98</td>
<td></td>
<td>Amer-Wahlin,2005</td>
</tr>
<tr>
<td>K: 0.31-1</td>
<td></td>
<td>Böss, 2003, 2005</td>
</tr>
<tr>
<td>Classification in categories</td>
<td></td>
<td>Chauhan, 2008</td>
</tr>
<tr>
<td>Normal</td>
<td>Moderate-good</td>
<td>Nielsen,87</td>
</tr>
<tr>
<td>Pa: 0.62</td>
<td></td>
<td>Lidegaard, 92</td>
</tr>
<tr>
<td>Susp/pathol</td>
<td>Poor-moderate</td>
<td>Amer-Wahlin,2005</td>
</tr>
<tr>
<td>Pa: 0.25-0.73</td>
<td></td>
<td>Böss, 2003, 2005</td>
</tr>
<tr>
<td>Overall</td>
<td>Poor-good</td>
<td>Chauhan, 2008</td>
</tr>
<tr>
<td>Pa: 0.18-0.89</td>
<td></td>
<td>Ojala,2008</td>
</tr>
<tr>
<td>K: 0.59</td>
<td></td>
<td>Waterlois,2009</td>
</tr>
<tr>
<td>wkK: 0.25-0.75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Legend: Accel – accelerations; decel – decelerations; susp – suspicious; pat – pathologic; Pa – proportion of agreement; K – kappa coefficient; wkK – weight kappa coefficient; ICC – intraclass correlation coefficient)
Estimation of the FHR baseline had a fair to good agreement in the majority of studies, while identification of accelerations and decelerations varied, and classification of decelerations constituted the CTG feature with the poorest agreement. Evaluation of FHR variability also showed a poor to moderate agreement across the studies.

Agreement on the overall classification of FHR tracings as normal was significantly better than that seen with suspicious or pathological categories. A positive correlation was observed between CTG interpretation, identification of FHR patterns and the number of years of clinical experience in some studies\textsuperscript{103,104}.

The available evidence on the poor inter- and intra-observer variability in FHR interpretation has put into question the influence that this phenomenon may have on the validity and effectiveness of the technique. It has also motivated further developments that include training programs, complementary techniques and eCTG analysis.

**CONTINUOUS $f$ECG ST WAVEFORM ANALYSIS**

As far back as 1957, it was recognized that ST segment depression and T-wave changes could be important markers of fetal distress\textsuperscript{105}. Many of the technical difficulties associated with the measurement of $f$ECG waveform characteristics were overcome over the next decades with the introduction of computer averaging, as a means to improve signal-to-noise ratio. Automatic computer-aided analysis became available during the 1980s with the development of the STAN® S-21 device (Neoventa, Moelndal, Sweden). By the year 2000 the STAN® methodology (Neoventa, Moelndal, Sweden) had become available for routine clinical use.

This system continuously monitors $f$ECG waveforms during labour by means of a scalp electrode. It calculates an “average $f$ECG complex” from 30 consecutively acquired ECG complexes, and analyses the former by calculating the T-wave amplitude (quantified by the T/QRS ratio) and classifying ST segment morphology into one normal and three abnormal classes (biphasic STs). A continuous read-out of T/QRS ratios is provided, together with additional information every time a biphasic ST is detected. The T/QRS ratio baseline is determined in order to identify clinically important T/QRS rises, and these are automatically flagged as ‘ST-events’. Likewise, clinically important biphasic ST segments (biphasic STs grade 2 or 3) of a repetitive nature are flagged as an “ST-event”\textsuperscript{105-107}.

The methodology requires an integrated interpretation of visual CTG analysis and automated ST waveform analysis. The CTG is classified according to guidelines that were developed specifically for this technology (STAN® guidelines), although inspired on those of the International Federation of Gynecology and Obstetrics (FIGO)\textsuperscript{108}.  

PhD THESIS | Antónia Costa | 2010
These guidelines also indicate, when an intervention is recommended. The technique has achieved widespread acceptance across mainland Europe and is also gaining popularity in other countries, including the United States of America.

**ECG PHYSIOLOGY**

Animal studies have shown that changes in the ST segment correspond to fetal reactions to hypoxia. The ST segment and T-wave represent the repolarisation phase of myocardial cells and this process is energy-consuming. An increase in T-wave amplitude occurs when the energy balance within myocardial cells threatens to become negative. This occurs when the amount of available oxygen is insufficient to cover energy requirements. The cells start to produce energy by β-adrenoceptor mediated anaerobic breakdown of glycogen reserves, a vital compensatory defense mechanism. This process not only produces lactic acid, but also releases potassium ions, which affect myocardial cell membrane potential and cause a rise in T wave amplitude. Thus, an increase in the T/QRS ratio reflects myocardial glycogenolysis and the use of a key defense mechanism in response to hypoxia.

ST depression with negative T waves has been observed in experimental settings during episodes of hypoxia. The explanatory physiology is a persistently negative energy balance causing a reduction in myocardial performance, which leads to an imbalance between endo- and epicardium. This indicates an incapacity to respond to hypoxia. However, others factors may also modify this balance within the myocardial wall and cause biphasic STs, such as prematurity, infections, temperature rise, marked tachycardia, metabolic imbalance and cardiac disturbances.

**SCIENTIFIC EVALUATION**

**VALIDITY**

It has been reported that ST events are present in 52% of all recordings in the first stage of labour and 24% in the second stage. Moreover, these events seem to have similar incidences during normal and abnormal FHR patterns in both stages of labour, but, when present, are more numerous in abnormal tracings.

An overview of the accuracy of combined CTG and ECG ST waveform analysis in prediction of adverse fetal outcomes is displayed in table IV.
Table IV: Accuracy of combined CTG and fECG ST waveform analysis

<table>
<thead>
<tr>
<th>Author, year</th>
<th>N</th>
<th>UA pH criteria (N)</th>
<th>Results</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amer-Wahlin, 2002</td>
<td>573</td>
<td>pH&lt;7.05 (10)</td>
<td>Sensitivity: 95.6%</td>
<td>▪ Lag time between end monitoring and birth not defined;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity: 96.4%</td>
<td>▪ Results include joint CTG and ST data;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PPV: 51%</td>
<td></td>
</tr>
<tr>
<td>Amer-Wahlin, 2005</td>
<td>142</td>
<td>pH&lt;7.05 (41) BD&lt;12mmol/l</td>
<td>Sensitivity: 63%</td>
<td>▪ Lag time between end monitoring and birth not defined;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity: 66%</td>
<td>▪ 11% of cases with low signal;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ Results include joint CTG and ST data;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ No analysis on ST-event type and the predictive capacity;</td>
</tr>
<tr>
<td>Derwaits, 2004</td>
<td>143</td>
<td>pH&lt;7.15 (7) BD≥12mmol/l</td>
<td>Sensitivity: 43%</td>
<td>▪ Lag time between end monitoring and birth not defined;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity: 74%</td>
<td>▪ Results include joint CTG and ST data;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PPV: 8%</td>
<td>▪ No analysis on ST-event type and the predictive capacity;</td>
</tr>
<tr>
<td>Kwee, 2004</td>
<td>449</td>
<td>pH&lt;7.05 (23) BD&lt;12mmol/l</td>
<td>Sensitivity: 65%</td>
<td>▪ Results include joint CTG and ST data;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity: 89%</td>
<td>▪ No analysis on ST-event type and the predictive capacity;</td>
</tr>
<tr>
<td>Luttkus, 2004</td>
<td>911</td>
<td>pH&lt;7.05 (53)</td>
<td>Sensitivity: 81%</td>
<td>▪ Lag time between end monitoring and birth not defined;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ 5 cases of acidosis inadequately recorded;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ No complete analysis on ST-event type and the predictive capacity;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ Results include joint CTG and ST data;</td>
</tr>
<tr>
<td>Norén, 2007</td>
<td>911</td>
<td>pH&lt;7.06-7.09 (44) BD&lt;12mmol/l</td>
<td>Sensitivity: 55%</td>
<td>▪ Lag time between end monitoring and birth not defined;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity: 77%</td>
<td>▪ 11% of cases with low signal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ Results include joint CTG and ST data;</td>
</tr>
<tr>
<td>Devoe, 2006</td>
<td>189</td>
<td>pH&lt;7.12 (16)</td>
<td>NPV: 95.2%</td>
<td>▪ No analysis on ST-event type;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ Results include joint CTG and ST data;</td>
</tr>
<tr>
<td>Vayssière,2007</td>
<td>411</td>
<td>pH≤7.05 (16)</td>
<td>Sensitivity: 62.5%</td>
<td>▪ No analysis on ST-event type;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity: 79%</td>
<td>▪ 38% of cases with low signal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PPV: 11%</td>
<td>▪ Results include joint CTG and ST data;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NPV: 98%</td>
<td>▪ No analysis on ST-event type and the predictive capacity;</td>
</tr>
<tr>
<td>Melin, 2008</td>
<td>506</td>
<td>pH≤7.15 (108)</td>
<td>ST event: 79%</td>
<td>▪ Up to 50% of cases with low signal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ Results include joint CTG and ST data;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ No analysis on ST-event type and the predictive capacity;</td>
</tr>
</tbody>
</table>

(legend: N – number of cases; UA – umbilical artery; BD – base deficit in extracellular fluid; NPV – negative predictive value; PPV – positive predictive value)

These observational studies assessed the predictive capacity of ST events or of the CTG+ST criteria used in the STAN® guidelines to recommend intervention, in the detection of abnormal UA pH and BD values. Different cut-off values for adverse
outcome were selected, thus somewhat limiting the comparison of results. Some studies report a high incidence of neonatal acidemia, which probably represents selection bias, and not all took into account the effect of training on the results.

From the results of these studies, sensitivities and specificities ranging between 62-96% for the detection of an UApH <7.05 were found, with negative and positive predictive values of 79-98% and 8-51%, respectively.

EFFECTIVENESS

Four randomised controlled trials have been published comparing the effect of combined CTG and fECG ST waveform analysis with isolated CTG, on the incidence of adverse neonatal and maternal outcomes. A fifth trial was recently concluded in the Netherlands, and its results are expected shortly. A summary of the main results of individual trials and a systematic review of the first three published trials is displayed in Table V.

Table V. Results of studies comparing combined CTG and fECG ST waveform analysis with isolated CTG.

<table>
<thead>
<tr>
<th>Authors, year</th>
<th>Total RCTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Westgate, 93</td>
<td>N=2434 N=4966</td>
</tr>
<tr>
<td>Amer-Wahlin, 2001</td>
<td></td>
</tr>
<tr>
<td>Ojala, 2006</td>
<td>N=1483</td>
</tr>
<tr>
<td>Vayssière, 2007</td>
<td>N=799</td>
</tr>
<tr>
<td>N=8883</td>
<td></td>
</tr>
</tbody>
</table>

**Outcomes**

- **Perinatal death**
  - Westgate, 93: 4.98 (0.24;103.70)
  - Amer-Wahlin, 2001: 1.46 (0.24;8.71)
  - Ojala, 2006: 0.33 (0.01;8.18)
  - Vayssière, 2007: 0.33 (0.11;0.95)
  - Total: 1.48 (0.42;5.24)

- **Neonatal encephalopathy**
  - Westgate, 93: 0.25 (0.03;2.23)
  - Amer-Wahlin, 2001: 0.36 (0.10;1.37)
  - Ojala, 2006: 0.34 (0.01;8.26)
  - Vayssière, 2007: NE
  - Total: 0.33 (0.11;0.95)

- **Metabolic acidosis**
  - Westgate, 93: 0.38 (0.14;1.07)
  - Amer-Wahlin, 2001: 0.47 (0.25;0.87)
  - Ojala, 2006: 1.21 (0.37;3.99)
  - Vayssière, 2007: 0.64 (0.25;1.63)
  - Total: 0.54 (0.35;0.83)

- **5-minute Apgar <7**
  - Westgate, 93: 0.62 (0.36;1.08)
  - Amer-Wahlin, 2001: 0.90 (0.53;1.53)
  - Ojala, 2006: 1.13 (0.44;2.92)
  - Vayssière, 2007: 1.00 (0.33;3.08)
  - Total: 0.81 (0.58;1.14)

- **Admission to NICU**
  - Westgate, 93: 0.77 (0.46;1.31)
  - Amer-Wahlin, 2001: 0.91 (0.74;1.11)
  - Ojala, 2006: 1.01 (0.59;1.72)
  - Vayssière, 2007: 0.84 (0.26;2.72)
  - Total: 0.90 (0.75;1.07)

- **CS rate**
  - Westgate, 93: 0.95 (0.74;1.21)
  - Amer-Wahlin, 2001: 0.92 (0.77;1.10)
  - Ojala, 2006: 1.35 (0.88;2.07)
  - Vayssière, 2007: 0.83 (0.60;1.16)
  - Total: 0.95 (0.84;1.08)

- **Instrumental birth**
  - Westgate, 93: 0.87 (0.74;1.02)
  - Amer-Wahlin, 2001: 0.85 (0.72;1.00)
  - Ojala, 2006: 0.89 (0.66;1.21)
  - Vayssière, 2007: 0.97 (0.73;1.27)
  - Total: 0.88 (0.79;0.97)

- **Operative deliveries**
  - Westgate, 93: 0.90 (0.79;1.01)
  - Amer-Wahlin, 2001: 0.88 (0.79;0.99)
  - Ojala, 2006: 1.03 (0.82;1.31)
  - Vayssière, 2007: 0.98 (0.86;1.11)
  - Total: 0.92 (0.86;0.98)

- **Operative delivery for fetal distress**
  - Westgate, 93: 0.55 (0.40;0.74)
  - Amer-Wahlin, 2001: 0.83 (0.69;0.99)
  - Ojala, 2006: 0.82 (0.57;1.16)
  - Vayssière, 2007: 0.91 (0.75;1.10)
  - Total: 0.79 (0.70;0.89)

**Legend**: N - number of cases; NN - neonatal; HI - hypoxic-ischaemic; NICU - neonatal intensive care unit; NE - not estimated; Metabolic acidosis defined as umbilical artery pH <7.05 and B D >12mmol/l; Results in relative risks (RR): result with 95% interval confidence (95IC) [a,b]; RR >1 favours CTG, RR<1 favours STAN®)
All of these studies have included efforts to obviate some of the methodological problems found in the early CTG studies. Standardised guidelines for FHR interpretation and management were used, training courses were carried out and acquisition of umbilical blood samples was uniformly accomplished (with the exception of sample processing).

It is important to refer that the results depicted on the table are an interpretation of the mentioned studies, the systematic review on the subject and some commentary articles on the same matter. In the study by Ojala et al. assessment of metabolic acidosis did not adopt the previously mentioned acid-base chart to estimate newborn acid-base status more accurately, and therefore an overestimation of the metabolic acidosis rate occurred. The corrected value is displayed in the table, as opposed to the one reported in the study and in the systematic review. Another important aspect of the study by Ojala et al. is that power calculation was not based on the incidence of the primary endpoint (UApH < 7.05), but rather on the incidence of UApH <7.10. As the incidence of pH<7.05 is lower, the study was underpowered to detect differences on this outcome. Vayssière et al conducted a small RCT of 799 participants with the following inclusion criteria: abnormal CTG or thick meconium-stained fluid, which, by itself, can constitute a selection bias. Besides STAN recording should be initiated as soon as possible in the intrapartum setting, since fetal heart may not behave as expected during periods of intense, but slowly progressive hypoxia and relevant ST features may not appear in this setting.

Overall, these results demonstrate that the addition of ECG ST waveform analysis to continuous CTG monitoring was associated with a decreased rate of metabolic acidosis at birth (defined as UApH <7.05 and BD >12mmol/l) and neonatal encephalopathy. It also reduced operative vaginal deliveries and total operative deliveries for fetal distress. There was no statistically significant effect on the incidence of low Apgar scores or admission to the NICU.

**REPRODUCIBILITY**

As referred previously, the aspect of the technology that relies on visual analysis of the CTG, may be associated with a limited interobserver agreement. Table VI displays the results of published studies that have evaluated the reproducibility of tracing interpretation and subsequent clinical decision using this methodology.
Table VI. Observer agreement on the STAN® methodology (Studies by author and year of publication; classification of FHR tracings is based on joint visual CTG and automated ST analysis)\textsuperscript{91,101,102,127-129}.

<table>
<thead>
<tr>
<th>Tracings classifications</th>
<th>Results</th>
<th>Agreement</th>
<th>Author, year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intraobserver K: 0.52-0.64</td>
<td>Moderate</td>
<td>Ojala, 2008</td>
<td></td>
</tr>
<tr>
<td>Interobserver K: 0.42-0.81</td>
<td>Moderate-good</td>
<td>Westerhuis, 2009</td>
<td></td>
</tr>
<tr>
<td>wK: 0.47-0.48</td>
<td>pa: 0.56-0.94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clinical decision-making</th>
<th>Tracings classifications</th>
<th>Results</th>
<th>Agreement</th>
<th>Author, year</th>
</tr>
</thead>
<tbody>
<tr>
<td>No intervention</td>
<td>Interobserver Pa: 0.76-0.94</td>
<td>Good</td>
<td>Ross, 2004</td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>Interobserver Pa: 0.50-0.89</td>
<td>Moderate-good</td>
<td>Amer-Wahlin, 2005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intraobserver K: 0.61-0.75</td>
<td>Moderate-good</td>
<td>DeVoe, 2006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>wK: 0.47-0.60</td>
<td>Pa: 0.80-0.96</td>
<td>Ojala, 2008</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>Interobserver Pa: 0.76-0.94</td>
<td>Good</td>
<td>Westerhuis, 2009</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vayssière, 2009</td>
<td></td>
</tr>
</tbody>
</table>

(legend: Pa – proportions of agreement; K – kappa coefficient; wK – weighted kappa coefficient)

As occurred with similar studies evaluating isolated CTG analysis, different measures of agreement were used, making comparison of results across the studies difficult. In general, agreement over the classification of tracings was moderate to good. Agreement on normal and pre-terminal traces appears to be better, which is a reassuring finding, since these patterns do not require additional ECG waveform information in order to decide management\textsuperscript{101}. The effect of observer experience on agreement is controversial. DeVoe et al reported similar results for experienced and non-experienced observers, while Westerhuis et al encountered a higher agreement in the less experienced and most experienced groups\textsuperscript{101,128}. The results in the less experienced group can perhaps be explained by the stricter application of the classification algorithms.

Agreement regarding clinical decision-making appears to be moderate to good with this methodology. On the other hand, in a study that analysed agreement on the timing of intervention, there was agreement within a time-frame of 30-60 minutes in only 26\% of cases and within 60-90 minutes in a further 23\%\textsuperscript{101}. In the face of what could be a rapidly deteriorating fetal state, this time frame seems somewhat excessive.

Interestingly, Ross et al presented FHR tracings to clinicians, first without ST data and subsequently with ST data, and concluded that the addition of ST data improved interobserver agreement on clinical decision and on timing of obstetric intervention\textsuperscript{128}. Similarly, the correct identification of the need to intervene in fetuses
with an adverse neonatal outcome increased from 86 to 93%, while unnecessary interventions decreased from 43 to 6% \cite{128}. Likewise, Amer-Wahlin et al. verified that interobserver agreement on abnormal CTG classification rose from 73 to 84%, when clinicians were given access to ST data \cite{50}. Interobserver rate on the indication to intervene also increased from 77 to 89% \cite{50}. Vayssière et al. concluded that in cases with abnormal CTG, ST analysis may improve consistency in clinical decision-making and decrease unnecessary interventions, but, on the other hand, access to ST data may also lead on rare occasions to unjustified decisions not to intervene \cite{129}. The rate of agreement for justified intervention was 94% (91-97%) with CTG and 85% (80-90%) with CTG+STAN (p< 0.001) and for justified nonintervention, 56% (48-63%) with CTG and 84% (79-89%) with CTG+STAN (p < 0.0001) \cite{129}.

Keeping a technology as simple as possible and eliminating problems with reproducibility are important aspects of wide dissemination and success, particularly when a large target audience, such as healthcare professionals managing intrapartum care, is involved \cite{130}. The development and validation of ST analysis, unlike continuous CTG monitoring, passed all the logical phases of the introduction of a new technique into clinical practice. This process included laboratory studies, evaluation of reproducibility, validity and effectiveness. The next expectable evaluation is the impact of the introduction of the technology on the quality of care in a given population. These studies have already been initiated and some have been published, showing a paradigm shift in the outcome of deliveries when a high rate of structured CTG+ST implementation is achieved in labour ward settings \cite{131-135}.

Despite these results, some adverse outcomes still occur with the use of fECG ST waveform analysis \cite{136}. A possible explanation for these is that the fetal heart may not behave as expected during periods of intense, but slowly progressive hypoxia and relevant ST features may not appear in this setting. On the other hand, reduced FHR variability seems to be a consistent finding in these cases \cite{137-139}. The awareness that the main weakness of this methodology lies in its dependence on visual CTG analysis has become widespread in the scientific community, suggesting that computer analysis may need to be incorporated in future developments \cite{137,140}.

**COMPUTERIZED CTG ANALYSIS**

The effectiveness of any diagnostic method is limited, not only by its ability to identify the adverse condition, but also by the ability of the clinical staff to interpret its results. As previously mentioned, numerous studies have shown that assessment of CTG tracings by healthcare professionals is prone to a high interobserver disagreement. This remains a major source of concern for the technology. Computerized analysis of CTGs arose as a method to overcome this problem, and several systems were
developed in different parts of the world. Much of this work remained in the experimental field and was not introduced in routine clinical practice. Moreover, i only a few systems was a detailed evaluation of validity carried out and published.

Basic technical concepts inherent to eCTG, such as sampling rate, signal loss, artifact detection, signal processing, digital archiving and monitor display have received considerable attention. The possible clinical and technical advantages of these systems are displayed in Table VII, together with the difficulties that can be encountered in their development.

**Table VII**: Advantages and limitations of systems for eCTG analysis systems.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Clinical</th>
<th>Technical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reproducibility</td>
<td></td>
<td>Reduced monitoring times in the antepartum\textsuperscript{141,142}</td>
</tr>
<tr>
<td>Systematic CTG analysis</td>
<td></td>
<td>Objective assessment of signal quality</td>
</tr>
<tr>
<td>Objective evaluation of CTG parameters unreliable quantified by human eye (variability)</td>
<td></td>
<td>Automatic digital storage and review</td>
</tr>
<tr>
<td>Easy access to real-time information in multiple posts</td>
<td></td>
<td>Automatic printing in low cost paper</td>
</tr>
<tr>
<td>On-line alerts</td>
<td></td>
<td>Automatic identification of tracings by time, date and place of acquisition</td>
</tr>
</tbody>
</table>

An overview of the main characteristics of systems for computer analysis of CTG tracings is presented in table VIII.
Many groups have concentrated on analysis of the more stable signal provided by antepartum CTG tracings, while others have designed systems specifically for intrapartum CTG analysis. Some systems may be used both for antepartum and intrapartum analysis. The increased complexity of intrapartum cCTG analysis is due to factors such as increased tracing length (and hence additional computer memory requirements), frequent occurrence of signal loss and artefacts, greater instability of the FHR baseline, greater clinical relevance of short-term variability and a higher number

![Table VIII: Main characteristics of cCTG systems (by author, country of main research)](image)

(Legend: Intra – intrapartum; Ante – antepartum; Clinical data – integration of obstetrical clinical data; Clinical alerts – clinical alerts displayed by the system based on abnormal FHR features for CTG classification versus clinical decision-making; NR – not reported in the studies)
of periodic events. The development of algorithms capable of dealing with these difficulties is a challenge to software development. Some systems analyse and classify CTG tracings offline, while others perform “real-time” analysis, which can be used to emit audio and visual alerts. A small number of groups have developed systems aimed at providing decision support for clinicians, and thus a suggested clinical management is displayed. The present thesis describes and evaluates the Omniview-SisPorto system for computer analysis of CTGs and eECG ST waveform analysis data, which will be considered in more detail in the following chapters.

REFERENCES


Chapter 2

Aim and outline of the thesis
## BRIEF OVERVIEW

Throughout the world, fetal monitoring during labour is performed with IA or continuous CTG recording. The latter has been increasingly used in industrialised countries, despite controversial scientific evidence in favour of its routine employment\(^1\). Interpretation of CTG tracings is generally performed by health professionals in charge of labour (labour and delivery nurses, midwives or obstetricians), but this has a well-demonstrated poor reproducibility\(^2-6\). Computer analysis of CTG tracings emerged as a way of overcoming this problem. Programs aimed at this objective have been developed over the last 3 decades. In spite of widespread interest and acclaimed need, eCTGs still plays a limited role in the clinical setting, especially in intrapartum care. There are many possible explanations for this. Firstly, there is still no consensus on which CTG parameters are best associated with fetal oxygenation, so different criteria are used by the systems. Secondly, the complexity of the intrapartum FHR signal constitutes a major challenge to the development of classification algorithms. Finally, this is a research area in which the clinical effectiveness of any methodology is very difficult to demonstrate, as has been shown by the several RCTs that evaluated intrapartum CTG\(^7-9\).

\(\delta\)ECG waveform analysis with automatic evaluation of the ST interval (STAN®, Neoventa, Moelndal, Sweden) has recently been added to intrapartum CTG monitoring. It was commercialised in the last decade of the 20\(^{th}\) century and spread rapidly throughout many industrialised countries. The system has been shown to decrease the incidence of newborn metabolic acidosis and HIE, while at the same time reducing operative delivery rates\(^10,11\). This methodology depends on the combined evaluation of CTG and \(\delta\)ECG data and is therefore influenced by the poor reproducibility of visual CTG interpretation\(^12-14\). Combining eCTGs with automatic evaluation of the ST interval has the potential to become a major developmental area in intrapartum fetal surveillance.

The Omniview-SisPorto® system (Speculum®, Lisbon, Portugal) was developed over the last 20 years at the University of Porto to perform computer analysis of CTG tracings. More recently it incorporated display and analysis of ST event data and the concepts of central monitoring of CTGs in multiple posts with real-time computer analysis of antepartum and intrapartum tracings\(^15-17\). The latter is used in order to trigger alerts for healthcare providers. It was commercialised in 2006 and has currently been installed in more than 20 hospitals worldwide.
AIM AND OUTLINE OF THE THESIS

The main objective of this thesis was to develop and evaluate a methodology that integrates computer cardiotocographic analysis with fetal electrocardiographic ST event data (as provided by the STAN® methodology), in order to improve the accuracy of intrapartum fetal monitoring in detection of abnormal fetal oxygenation. The ultimate objective is to increase the effectiveness of intrapartum fetal monitoring by reducing the incidence of adverse neonatal outcomes without increasing the number of operative deliveries. In order to accomplish this objective, the following research plan phases were established:

1. Optimisation of the Omniview-SisPorto® system and development of algorithms for joint cardiotocographic and electrocardiographic ST event analysis in order to provide real-time clinical alerts. Chapter 3

2. Comparison of the system’s analysis of basic cardiotocographic features, baseline, accelerations, decelerations and contractions with that of a consensus of experts. Chapter 4

3. Evaluation of the impact of access to computerised cardiotocogram analysis on clinicians’ reproducibility and accuracy in prediction of neonatal outcome parameters. Chapter 5

4. Evaluation of the accuracy of the system’s alerts in prediction of umbilical artery acidemia. Chapter 6

5. Planning of the assessment of the system’s effectiveness in routine clinical practice by designing a randomised controlled trial to evaluate its effect on neonatal and maternal outcomes. Chapter 7

REFERENCES


Omniview-SisPorto® 3.5 - a central fetal monitoring station with on-line alerts based on computerised cardiotocogram+ST events signals.

Diogo Ayres-de-Campos¹,², Paulo Sousa², Antónia Costa¹,², João Bernardes¹,²

¹ Department of Obstetrics and Gynaecology, Faculty of Medicine, Porto University, São João Hospital, Portugal
² Institute of Biomedical Engineering, Porto
ABSTRACT

Visual analysis of cardiotocograms is poorly reproducible and is currently recognised as the main weakness of the STAN® methodology. The Omniview-SisPorto® 3.5 program is the most recent version of a central monitoring system that provides visual and sound alerts, based on computer analysis of cardiotocographic and ST event features. This paper describes the program’s main characteristics and provides an overview of the system’s online alerts. Omniview-SisPorto® 3.5 is the first central monitoring system to incorporate computerised analysis of cardiotocographic and ST event features, providing health professionals with online alerts for minor and major changes in monitored signals. The system is currently undergoing extensive clinical evaluation.
INTRODUCTION

Intrapartum fetal monitoring with combined evaluation of cardiotocographic and electrocardiographic signals forms the basis of the STAN® technology, which has been shown in randomized trials to decrease the incidence of HIE and instrumental deliveries, when compared to CTG alone⁴. While providing automated evaluation of ECG features, the former technique still relies heavily on human interpretation of CTG tracings for clinical decision. It has been well demonstrated that visual analysis of CTGs has a limited reproducibility², and this is currently recognised as the main weakness of the STAN® methodology³. Computer systems for CTG analysis provide a totally reproducible evaluation of the signal and allow the quantification of parameters that cannot be reliably assessed by the human eye, such as short- and long-term variability. They have been developed since the late 1970s, but because of the difficulties caused by signal instability in intrapartum tracings, have mainly focused on the antepartum period.

This paper gives a brief description of the Omniview-SisPorto® 3.5 system (Speculum, Lisbon, Portugal), the latest version of a program for automated analysis of CTGs developed over the last 19 years at the University of Porto and provides an overview of the system’s online alerts⁴,⁵.

DESCRIPTION OF THE SYSTEM

The Omniview-SisPorto® 3.5 system was designed as a central monitoring station, allowing the combined display of up to 16 CTG and/or STAN® tracings on the same computer screen, which constitutes the main program window (Figure 1).
**Figure 1.** *(Above)* Main program window showing simultaneous display of three tracings and online alerts. *(Below)* Maximised tracing window, showing detailed identification of FHR baseline, accelerations and decelerations with a combined CTG+ST online alert displayed underneath the tracing.
Computer processing of CTG features starts automatically after 11 minutes of tracing acquisition and is updated every minute. This includes estimation of FHR baseline, identification of accelerations and decelerations and quantification of short- and long-term variability, according to previously described algorithms. A detailed display of the results can be accessed by maximizing a single tracing window. A combination of CTG analysis with ST event features is used to provide online alerts, which are displayed in the main program window, as well as in maximised tracing windows.

Tracing acquisition, display and storage is handled automatically by the program, as it continuously detects the connection and disconnection of fetal monitors transmitting appropriate signal features. External input is only required for identification of patients in a local or external database. A multiclcent-server architecture ensures that a practically unlimited number of computers can be connected via a local intranet to display tracings. Exams can be printed automatically on tracing end or at the user’s request, with or without the results of automated analysis. Previously acquired tracings can be processed as a whole or analysed as if they were being acquired online, at normal or higher speeds.

**ONLINE ALERTS**

Processing of information to produce online alerts starts automatically after 11 minutes of acquisition and is updated every minute. Definitions used for alerts were inspired on published consensus guidelines for interpretation of CTGs and combined CTG+ST signals, but were adapted according to the results of previous optimisation and validation studies. Alerts are divided into six categories (Table 1).

“Attention alerts” occur with changes, that are not commonly associated with fetal hypoxia, such as CTG and ST signal loss, excessive uterine contractions, and maternal heart rate monitoring. They produce a blue visual alert with no sound.

“Attention extra alerts” occur with changes that may or may not be associated with fetal hypoxia, such as large signal loss or a CTG pattern compatible with both maternal heart rate monitoring and a FHR deceleration. They produce a blue blinking visual alert and a configurable sound alert, calling the attention of health professionals for the possible need to act on the situation.

The “Normality criteria met alert” was designed for short-lasting CTG tracings, such as those that are acquired in outpatient clinics or as labour admission tests. It is only processed between 20 and 60 minutes after start of tracing acquisition, when all normality criteria regarding FHR baseline, accelerations, decelerations and variability are
present. It produces a green alert with no sound and may be excluded in the configurations.

“Non-reassuring alerts” occur with changes in the FHR that are usually not associated with fetal hypoxia, but may require continued monitoring and occasionally further investigation: basal bradycardia, FHR baseline shift, fetal tachycardia, absent accelerations, infrequent non-prolonged decelerations or repetitive but short-lasting decelerations. They produce a yellow alert with no sound.

“Very non-reassuring alerts” occur with more worrying changes in the FHR, which are usually associated with a considerable risk of impending fetal hypoxia: repetitive and moderate-lasting decelerations, prolonged decelerations or repetitive short-lasting decelerations with single mild ST event. Careful monitoring and further measures are usually required to avoid progress to relevant degrees of hypoxia. They produce a blinking orange visual alert with sound.

“Pre-terminal alerts” occur with signal changes that are commonly associated with moderate-to-severe fetal hypoxia: reduced long-term and short-term variability, repetitive or prolonged decelerations with reduced variability or ST events, tachycardia, decelerations and ST event or very prolonged deceleration. Rapid measures are required in these cases to avoid an abnormal fetal outcome, which may include rapid reversal of the underlying cause or immediate delivery.
Table 1. Description of Omniview-SisPorto® 3.5 online alerts.

<table>
<thead>
<tr>
<th>ATTENTION</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal loss</td>
<td>signal loss &gt;15% in last 10 min</td>
</tr>
<tr>
<td>Excessive uterine contractions</td>
<td>&gt;5 contractions in last 10-11 min or &gt;7 contractions in last 15 min (valid for a period of 20 min)</td>
</tr>
<tr>
<td>Maternal heart rate monitoring?</td>
<td>current signal features compatible with maternal heart rate</td>
</tr>
<tr>
<td>ST signal loss</td>
<td>no ST data in last 4 min</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ATTENTION EXTRA (configurable sound)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Large signal loss</td>
<td>signal loss &gt;50% in last 5 min</td>
</tr>
<tr>
<td>Maternal heart rate monitoring?</td>
<td>current signal compatible with maternal heart rate or deceleration lasting 2-8 min</td>
</tr>
<tr>
<td>Decelerations?</td>
<td>current signal compatible with maternal heart rate or deceleration lasting &gt;8 min</td>
</tr>
<tr>
<td>Maternal heart rate monitoring?</td>
<td></td>
</tr>
<tr>
<td>very prolonged deceleration?</td>
<td></td>
</tr>
</tbody>
</table>

| REASSURING (only occurs between first 20-60 minutes when all characteristics are present) | |
| Normality criteria met | (baseline and baseline FHR 105-160 bpm) and (≥2 accelerations in last 50 min) and (aLTV <10% in last 30 min or 10 consecutive minutes of absent aLTV in last 50 min) and (≤1 deceleration < 2 min in last 50 min) and no non-reassuring or pre-terminal alarms |

<table>
<thead>
<tr>
<th>NON-REASSURING</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Basal bradycardia</td>
<td>basal FHR &lt; 105 bpm</td>
</tr>
<tr>
<td>Baseline shift</td>
<td>FHR baseline &gt; 20 bpm than basal FHR in last 10 min</td>
</tr>
<tr>
<td>Fetal tachycardia</td>
<td>basal FHR &gt; 160 bpm or baseline FHR &gt; 160 bpm in last 20 min</td>
</tr>
<tr>
<td>No accelerations</td>
<td>≤1 acceleration in last 50 min</td>
</tr>
<tr>
<td>Decelerations</td>
<td>≥1 decelerations lasting 2-5 min in last 5 min</td>
</tr>
<tr>
<td>Repetitive decelerations</td>
<td>≥2 decelerations in last 10 min and ≥4 decelerations in last 20 min</td>
</tr>
</tbody>
</table>

| VERY NON-REASSURING (sound) | |
| Very repetitive decelerations | [(≥3 decelerations in last 10 minutes and ≥7 decelerations in last 20 min) and (≥1 deceleration lasting > 1 min in last 20 min)] or [(≥2 decelerations in last 10 min and ≥4 decelerations in last 20 min) and (≥1 deceleration > 2 min in last 20 min)] |
| Prolonged deceleration: | ≥1 deceleration lasting 5-8 min in last 10 min with normal STV in last 5 minutes |
| Repetitive decelerations and ST event | (repetitive decelerations) and (1 biphasic ST or 1 episodic T/QRS rise 0.10-0.15 or 1 baseline T/QRS rise 0.05-0.10), all in the last 20 min |

| PRETERMINAL (sound) | |
| Reduced long-term variability: | aLTV > 31% in the last 10 and 50 min |
| Repetitive decelerations and reduced long-term variability | repetitive decelerations and aLTV > 31% in the last 20 min |
| Basal bradycardia, decelerations and ST event | (baseline > 160 bpm) and (decelerations lasting >2 min or repetitive decelerations) and (≥1 of episodic T/QRS rise or ≥1 baseline T/QRS rise), all in the last 20 min |
| Repetitive decelerations and ST event | (repetitive decelerations in last 20 min) and (≥1 biphasic ST or ≥1 episodic T/QRS rise > 0.15 or ≥1 baseline T/QRS rise > 0.10) in the last 10 min |
| Decelerations and ST event | (decelerations lasting 2-5 min) and (≥1 biphasic ST or ≥1 episodic T/QRS rise > 0.15 or ≥1 baseline T/QRS rise > 0.10), all in the last 20 min |
| Very repetitive decelerations and ST event | (very repetitive decelerations) and (≥1 biphasic ST or ≥1 episodic T/QRS rise or ≥1 baseline T/QRS rise), all in the last 20 min |
| Prolonged decelerations and ST event | (decelerations lasting >5 min) and (≥1 biphasic ST or ≥1 episodic T/QRS rise or ≥1 baseline T/QRS rise), all in the last 5 min |

| Reduced short-term variability | |
|-------------------------------|
| Real beat/beat signals: aSTV > 92% in last 10 and 20 min |
| ECG with 4 Hz sampling: aSTV > 68% in last 10 and 20 min |
| US with 4 Hz sampling: aSTV > 92% in last 10 and 20 min |
| Very repetitive decelerations and absence of “spiky” accelerations in last 30 min and: | |
| Real beat/beat signals: aSTV > 65% in last 10 or 20 min |
| ECG with 4 Hz sampling: aSTV > 89% in last 10 or 20 min |
| US with 4 Hz sampling: aSTV > 65% in last 10 or 20 min |

| Repetitive decelerations and reduced short-term variability | |
|-----------------------------------------------|
| Real beat/beat: aSTV > 65% in last 5 min |
| ECG with 4 Hz sampling: aSTV > 80% in last 5 min |
| US with 4 Hz sampling: aSTV > 85% in last 5 min |

| Very prolonged deceleration | |
|----------------------------|
| deceleration lasting > 8 min in last 5 min |
| deceleration > 5 min in last 5 min and: | |
| Real beat/beat: aSTV > 65% in last 5 min |
| ECG with 4 Hz sampling: aSTV > 80% in last 5 min |
| US with 4 Hz sampling: aSTV > 85% in last 5 min |

Legend: aSTV – abnormal short-term variability (subsequent FHR points differing less than 1bpm); aLTV – abnormal long-term variability (FHR points with difference between minimum and maximum values in surrounding 1-minute window <5bpm).
**DISCUSSION**

The Omniview-SisPorto® 3.5 system was designed to incorporate the concepts of central monitoring, computerised CTG analysis and online alerts, using combined CTG + ST event data. The main aim of this technology is to alert health professionals to situations that may be associated with fetal hypoxia, thus preventing that such occurrences are overlooked in busy labour ward settings. It is also intended as an adjunct to clinical decision, especially regarding CTG changes that are poorly assessed by the human eye or when complex integration of CTG and ST information is needed. It is hoped that this technology, when associated with appropriate training of health professionals on signal interpretation, clinical decision and adequate management, will contribute to reduce perinatal morbidity and mortality consequent to intrapartum events.

Omniview-SisPorto 3.5 has been installed in over 15 European hospitals and has now been tested in more than 20,000 pregnancies. It is the first central monitoring system to incorporate computerised analysis of CTG and ST event features, providing health professionals with online visual and sound alerts. This version of the system is currently undergoing extensive clinical evaluation.

**REFERENCES**

Comparision of the evaluation of cardiotocographic events by Omniview-SisPorto® 3.5 and a consensus of clinicians.

Antónia Costa¹², Diogo Ayres-de-Campos¹², Ana Paula Machado³, Cristina Santos⁴⁵, João Bernades¹²

¹ Department of Obstetrics and Gynaecology, Faculty of Medicine, Porto University, São João Hospital, Portugal
² Institute of Biomedical Engineering, Porto
³ Department of Obstetrics and Gynaecology, São João Hospital, Portugal
⁴ Department of Biostatistics and Medical Informatics, Faculty of Medicine, Porto University
⁵ Centre for Research in Health Technologies and Information Systems, Faculty of Medicine, Porto University
ABSTRACT

**Aims** – The objective was to compare computer analysis of intrapartum CTG features by Omniview-SisPorto® 3.5 and a consensus of clinicians.

**Methods** – Agreement study using 50 consecutively acquired tracings (206 hours of signals) with more than 60 minutes duration, less than 10% signal loss and recorded in labour at term by internal fetal heart rate monitoring. Tracings were divided into ten-minute segments and independently analysed by three experienced clinicians in order to estimate the FHR baseline and identify periodic events. A consensus was reached using a three round *Delphi* procedure. Results were compared with the analysis provided by the Omniview-SisPorto® 3.5 system.

**Results** – For baseline estimation, agreement between the computer and the consensus was high (ICC: 0.85 [0.46;0.93]95CI, with a mean difference of 3.7 bpm (LA: -4.4 to 11.9 bpm), and 99% of differences under 15 bpm. A concordant identification was observed in 71% [0.69;0.73]95CI of accelerations, 68% [0.66;0.70]95CI of decelerations and 87% [0.85;0.89]95CI of uterine contractions.

**Conclusions** – A high agreement was observed between Omniview-SisPorto® 3.5 and a consensus of clinicians in evaluation of intrapartum CTG baseline, accelerations, decelerations and uterine contractions.
INTRODUCTION

Continuous cardiotocographic monitoring during labour is widely employed in industrialised countries, despite controversial scientific evidence regarding its benefits\(^1\). Analysis of CTG tracings by health professionals is almost universally performed visually, but this has a well-demonstrated poor reproducibility, both in overall classification\(^2,3\) and in identification of basic CTG events\(^4,5\). Systems for eCTGs have emerged over the last three decades as a way of overcoming this problem. Antepartum CTG has constituted the main area of research and development\(^6,7\), while several limitations have been reported, when these systems are applied during labour\(^7,8\).

Intrapartum tracings pose additional difficulties for computer analysis, related to increased tracing length (and hence additional computer memory requirements), higher signal loss, artefacts and signal instability. The latter resulting in increased difficulties for baseline estimation.

Omniview-SisPorto\(^®\) 3.5 (Speculum, Lisbon, Portugal)\(^9\) is the latest version of a system for computer analysis of CTG tracings, developed over the last 19 years at the University of Porto\(^10,11\). It incorporates the concepts of centralised CTG monitoring with multiple viewing posts and computer analysis with real-time alerts, based on joint evaluation of CTG and ST event features\(^9\). Computer analysis of CTGs follows the classical steps used in visual analysis: baseline estimation, followed by identification of accelerations, decelerations, contractions and finally evaluation of short- and long-term variability. The program has been adapted for analysis of both antepartum and intrapartum tracings.

While computerised analysis of CTGs was developed to overcome the limited reproducibility of visual interpretation and thus to develop an objective form of identifying fetuses undergoing relevant hypoxia, it is important that healthcare professionals relate with the basic aspects of tracing interpretation by the computer. This is particularly true of a system, which relies on the correct evaluation of the classical features of CTG analysis, such as baseline, accelerations, decelerations and contractions to elicit clinical alerts for normal and abnormal CTG patterns. The aim of this study was to compare the analysis of intrapartum cardiotocographic events (baseline, accelerations, decelerations and contractions) performed by Omniview-SisPorto\(^®\) 3.5 system with a similar evaluation undertaken by a consensus of experienced clinicians.

MATERIAL AND METHODS

A database of tracings acquired during a previously reported RCT\(^12\) was searched in order to select the first 50 cases with more than 60 minutes duration and less than 10% signal loss. All tracings had been recorded during labour in singleton pregnancies with
more than 36 weeks gestation by internal FHR monitoring and external uterine activity
detection, using the STAN® 21 fetal monitor (Neoventa Medical, Gothenburg,
Sweden). The selected tracings corresponded to a total of 206 hours of recording with
an average tracing length of 4 hours and 15 minutes and a maximum length of 16
hours.

Tracings were printed on paper at a 1 cm/min scale, using the Omniview-SisPorto®
3.5 software, but with no computer analysis. They were manually divided using vertical
lines into 10-minute segments. Three copies were prepared and sent for independent
visual analysis to three clinicians, all of whom had more than five years everyday clinical
experience in CTG interpretation and were not involved in the development of the
Omniview-SisPorto® 3.5 algorithms (AC, APM, JB). All observers considered
computer-printed recordings to be indistinguishable from CTG-paper printed tracings
and used a paper speed of 1cm/min regularly in their clinical practice. Clinicians were
only informed of the gestational age at which the tracing was acquired and the signal
acquisition method.

The three clinicians were asked to estimate the FHR baseline in each 10-minute
segment and then to identify accelerations, decelerations and uterine contractions. For
baseline estimation a previously developed definition was used, which has been shown
to be very reproducible13: “it is a single value, corresponding to the mean FHR of the lowest stable
horizontal segment(s) lasting at least 2 minutes. For the selection of these segments the following
conditions should preferably be met: long-term variability less than 15 bpm, absence of fetal movements
and uterine contractions and mean FHR within physiological limits”. If baseline estimation in a
particular 10-min segment was considered particularly difficult, observers were allowed
to consider the preceding segment(s) to arrive at their estimation, but not the
subsequent ones. Accelerations and decelerations were defined according to the RCOG
guidelines14, while no definition of uterine contractions (UC) was provided.

The first round of analysis was carried out independently by the three clinicians.
Baseline values were considered discordant, if they differed by more than 5 bpm. A
maximum tolerance window of 30 seconds on each side was allowed for the
identification of accelerations, decelerations and UC. A second round of independent
analysis was promoted in discordant segments, but with no information of other
observers’ evaluations.

For the third round of agreement a consensus meeting with the three clinicians and
an arbitrator (DAC) was convened, where all second round discordant segments were
reviewed. Because of the high number of segments needing re-evaluation, this took
place over three different sessions for a total of six hours. The events, in which
consensus was not reached in this round, were excluded from further analysis.
All tracings were then submitted to computerised analysis by the Omniview-SisPorto® 3.5 system, using an available routine that simulates online acquisition: i.e. computer analysis starts at the end of the first ten minutes of the recording and is subsequently updated every minute, but does not incorporate subsequently available signals into the analysis.

For each 10-minute segment the baseline attributed by Omniview-SisPorto® 3.5 was compared with that estimated by the consensus of clinicians and a similar methodology was applied for the identification of accelerations, decelerations and contractions; allowing a maximum 30-second tolerance window on each side of the event.

STATISTICAL METHODS

For evaluating agreement in baseline estimation, the ICC, the Pa and the LA were used\textsuperscript{15,16}. For identification of accelerations, decelerations and contractions, agreement was assessed using the proportions of specific agreement (Psa) calculated with $95\%$CI\textsuperscript{17,18}. Values of the ICC exceeding 0.75 were interpreted as corresponding to a high agreement\textsuperscript{19}. As proposed by Grant et al, if the $95\%$CI of the Pa included 0.5, then agreement was considered poor\textsuperscript{20}. When the values of the LA represented clinically acceptable differences between methods, agreement was considered adequate\textsuperscript{15}. Statistical methods were calculated using Microsoft\textsuperscript{®} Excel 2003 and SPSS for Windows\textsuperscript{®} version 10.0.7.

RESULTS

A total of 1293 ten-minute segments were obtained after tracing segmentation, of which 47 were excluded (3.6\%), because of short duration of the FHR signal or excessive signal loss during that segment, resulting in a final number of 1246 segments submitted for analysis. The three clinicians identified a mean number of 2643 accelerations, 1327 decelerations and 1853 UC in these segments. Omniview-SisPorto® 3.5 identified 2895 accelerations, 1476 decelerations and 1915 contractions.

After the first round there was a difference in baseline estimation exceeding 5 bpm between at least two observers in 438 segments. After the second round this number decreased to 39 and in the last round consensus was reached in all but two segments.

Concerning identification of accelerations, decelerations and UC, after the first round there were 1529 segments, where disagreements occurred (839 on accelerations, 549 on decelerations and 141 on UC). After the second round this number decreased to
323 (178 on accelerations, 143 on decelerations and 2 on UC). Consensus was reached in all but 27 segments (seven on accelerations and 20 on decelerations).

Table 1 displays the results of the evaluations of agreement on baseline estimation that were performed.

**Table 1:** Agreement on baseline estimation (between observers’ first-round analysis, between individual observers’ first-round analysis and Omniview-SisPorto® 3.5, between individual observers’ first-round analysis and their consensus and between Omniview-SisPorto® 3.5 and the consensus)

<table>
<thead>
<tr>
<th></th>
<th>ICC [95CI]</th>
<th>Mean Difference (bpm)</th>
<th>LA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observers 1,2 and 3</td>
<td>0.87 [0.84;0.90]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observer 1 and computer</td>
<td>0.79 [0.48;0.89]</td>
<td>4.01 [-6.59;14.61]</td>
<td></td>
</tr>
<tr>
<td>Observer 2 and computer</td>
<td>0.88 [0.74;0.93]</td>
<td>2.44 [-5.60;10.49]</td>
<td></td>
</tr>
<tr>
<td>Observer 3 and computer</td>
<td>0.78 [0.27;0.91]</td>
<td>5.05 [-5.25;15.35]</td>
<td></td>
</tr>
<tr>
<td>Observer 1 and consensus</td>
<td>0.93 [0.92;0.94]</td>
<td>0.29 [-6.90;7.48]</td>
<td></td>
</tr>
<tr>
<td>Observer 2 and consensus</td>
<td>0.95 [0.91;0.97]</td>
<td>-1.27 [-6.82;4.28]</td>
<td></td>
</tr>
<tr>
<td>Observer 3 and consensus</td>
<td>0.92 [0.90;0.94]</td>
<td>1.23 [-6.24;8.91]</td>
<td></td>
</tr>
<tr>
<td>Consensus and computer</td>
<td>0.85 [0.46;0.93]</td>
<td>3.71 [-4.44;11.87]</td>
<td></td>
</tr>
</tbody>
</table>

(Legend: ICC – intraclass correlation coefficient; 95CI - 95% confidence intervals; LA - Bland and Altman limits of agreement; consensus – consensus of clinicians).

High levels of agreement (ICC >0.75) were obtained in all evaluations. In 67% of segments differences in baseline estimation between Omniview-SisPorto® 3.5 and the consensus were less than 5 bpm, in 94% they were less than 10 bpm and in 99% they did not exceed 15 bpm. The maximum difference in baseline estimation observed between Omniview-SisPorto® 3.5 and the consensus was 21 bpm (Figure 1).
Figure 1. Plot graph of the differences in baseline estimation between Omniview-SisPorto® 3.5 and the consensus.

The results of the agreement evaluations on identification of accelerations, decelerations and contractions are displayed in Table 2. A satisfactory agreement (≥ 68%) was observed between Omniview-SisPorto® 3.5 and the consensus in identification of all parameters.

Table 2: Agreement on identification of accelerations, decelerations and contractions: number of agreements/ total number of identified features [percentage of agreements (95CI in brackets)].

<table>
<thead>
<tr>
<th></th>
<th>Accelerations</th>
<th>Decelerations</th>
<th>Contractions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observers 1, 2 and 3</td>
<td>1586/2643 [60% (48;66)]</td>
<td>1053/1620 [65% (57;69)]</td>
<td>1711/1840 [93% (90;95)]</td>
</tr>
<tr>
<td>Observer 1 and computer</td>
<td>2157/3172 [68% (52;75)]</td>
<td>1028/1631 [63% (51;68)]</td>
<td>1779/2069 [86% (83;88)]</td>
</tr>
<tr>
<td>Observer 2 and computer</td>
<td>2189/3172 [69% (55;76)]</td>
<td>1011/1631 [62% (49;65)]</td>
<td>1738/2069 [84% (83;87)]</td>
</tr>
<tr>
<td>Observer 3 and computer</td>
<td>2062/3172 [65% (50;71)]</td>
<td>995/1631 [61% (51;68)]</td>
<td>1759/2069 [85% (81;90)]</td>
</tr>
<tr>
<td>Observer 1 and consensus</td>
<td>1877/2643 [71% (63;82)]</td>
<td>1134/1620 [70% (59;76)]</td>
<td>1711/1840 [93% (90;97)]</td>
</tr>
<tr>
<td>Observer 2 and consensus</td>
<td>1850/2643 [70% (65;77)]</td>
<td>1118/1620 [69% (59;77)]</td>
<td>1711/1840 [93% (87;95)]</td>
</tr>
<tr>
<td>Observer 3 and consensus</td>
<td>1850/2643 [70% (63;78)]</td>
<td>1118/1620 [69% (61;81)]</td>
<td>1693/1840 [92% (88;96)]</td>
</tr>
<tr>
<td>Consensus and computer</td>
<td>2252/3172 [71% (69;73)]</td>
<td>1110/1631 [68% (66;70)]</td>
<td>1800/2069 [87% (85;89)]</td>
</tr>
</tbody>
</table>
DISCUSSION

A high agreement was found between Omniview-SisPorto® 3.5 and a consensus of clinicians in baseline estimation and in identification of accelerations, decelerations and contractions.

With the exception of UC, agreement was either similar or significantly higher to interobserver agreement and to agreement between individual observers and the consensus. This is a relevant finding as computer assignment is particularly difficult during the intrapartum, especially when small window segments are analysed.

Agreement on baseline estimation is particularly important, as this step can have a profound influence on the remaining CTG analysis, and it is one of the most reproducible aspects of visual CTG evaluation. The results obtained suggest, that satisfactory baseline estimation is provided by the Omniview-SisPorto® 3.5 system, but that further improvements can still be made, as clinicians estimated a value on average 3.71 bpm higher and the 95% upper range of the limits of agreement is a little under 12 bpm.

A previous evaluation of baseline estimation was performed using an earlier version of the Omniview-SisPorto® system that compared a single-value estimation over 40-60 minute tracings with a consensus of three clinicians. In the 150 intrapartum cases evaluated in that study, the results for the kappa statistic were 0.87 for the Pa and 0.95 for ICC (allowing a 5 bpm maximum difference between estimations). The LA were -6.45 to 7.07 (mean difference of 0.13 and standard deviation (sd) of 3.45). Clearly the evaluation of the Omniview-SisPorto® 3.5 wandering baseline in 10-minute segments, as performed in the current study, is more challenging.

Studies, that evaluate computerised FHR baseline estimation during labour, using other systems, are rare. Mongelli et al evaluated agreement on baseline estimation between 12 experts and an algorithm developed from the Nottingham/Hong-Kong system in 60 intrapartum tracings. Comparing each observer with the system, they reported ICC ranging from 0.83 to 0.97. Ninety-five percent of the differences were situated between –12 and +15 bpm and the maximum difference exceeded 50 bpm. The computer was unable to determine the baseline in 4 tracings.

Devoe et al compared visual analysis of 50 intrapartum tracings by four observers, according to the National Institute of Child Health and Human Development guidelines with each other and with those of an automated FHR monitoring system (TraceVue, Philips Medical, Germany). For baseline estimation, allowing a plus or minus window of 5 bpm, there was a 98-99% agreement between observers and an 84-88% agreement between computer and observers. Regarding quantification of accelerations, agreement between observers was 50-62% and between
computer and observers was 44-62%. In quantification of decelerations, agreement between observers was 36-51% and between computer and observers was 43-67%.

The interpretation of the agreement on identification of accelerations and decelerations is much more uncertain than that of the baseline, due to the higher interobserver disagreement found in evaluation of these parameters. Nevertheless, it is reassuring to observe that the computer agrees with the consensus analysis as much as do individual clinicians. The results obtained in the present study suggest a higher agreement in identification of accelerations and decelerations to that obtained with other systems\textsuperscript{23}. It became clear during data analysis, that periodic events with an amplitude and/or duration close to the cut-off values were the most frequent cause of disagreement. This occurred both in interobserver analysis and in the computer-consensus comparison. Agreement over identification of periodic events is particularly sensitive to small disagreements in baseline estimation, which can by itself lead to divergence over the former. The clinical significance of disagreement over identification of accelerations and decelerations of such small amplitude is probably very limited at term.

To our knowledge, no other studies have compared computer systems with clinicians on identification of UC. Our results suggest that a high interobserver agreement can be obtained in identification of this parameter and that Omniview-SisPorto\textsuperscript{®} 3.5 provides an acceptable evaluation, although further improvements in this specific algorithm could probably be made in order to allow a similar agreement to that obtained between observers.

While the main motive for development of systems for computer analysis of CTGs is the well known poor agreement on visual interpretation, the latter results in the absence of a gold standard with which to compare computer analysis. It could be argued that such comparison is unnecessary, as long as computer systems reliably predict neonatal outcome. But an accurate prediction of these outcomes relies on an adequate identification of basic CTG features, such as FHR baseline, accelerations, decelerations and UC. Analysis of variability is another important aspect for CTG interpretation, but agreement on this parameter was purposely left out of this study, as it has become generally accepted that computers are better tools at quantifying this\textsuperscript{24,25}. Evidence of poor interobserver agreement on visual estimation of variability is abundant, even when dichotomic assessment (normal \textit{versus} abnormal) is solicited\textsuperscript{4,5,24-26}. Kappa values of 0.13-0.35 have been reported, when this parameter is assessed in intrapartum tracings\textsuperscript{5} and 0.38 when assessed in antepartum tracings\textsuperscript{24}. These findings suggest that there is limited value in using visual analysis as a gold standard for comparing computerised quantification of variability.

We believe that for the successful introduction of computerised systems for CTG analysis into clinical practice, healthcare professionals need to understand the main
reasons behind computer-derived alerts and feel that these relate with the basic features of tracing interpretation. Thus, even if a high accuracy of alerts is achieved in identifying fetuses with relevant hypoxia, if it is not known how these alerts are elicited and/or if health professionals feel that they are in disagreement with basic tracing interpretation, they may look still at these systems with scepticism or even decide to ignore their alerts.

REFERENCES


Access to computerised analysis of intrapartum cardiocotographs improves clinician’s prediction of newborn umbilical pH.

Antónia Costa\textsuperscript{1,2}, Cristina Santos\textsuperscript{3,4},
Diogo Ayres-de-Campos\textsuperscript{1,2}, Célia Costa\textsuperscript{1,2}, João Bernardes\textsuperscript{1,2}

\textsuperscript{1} Department of Obstetrics and Gynaecology, Faculty of Medicine, Porto University, São João Hospital, Portugal
\textsuperscript{2} Institute of Biomedical Engineering, Porto
\textsuperscript{3} Department of Biostatistics and Medical Informatics, Faculty of Medicine, Porto University
\textsuperscript{4} Centre for Research in Health Technologies and Information Systems, Faculty of Medicine, Porto University
ABSTRACT

Objective – To evaluate the impact of access to eCTG analysis on clinicians’ reproducibility and accuracy in prediction of UApH and 5-minute Apgar score.

Design – Prospective evaluation of pre-recorded cases.

Setting – A tertiary care university hospital.

Population – From a databases of intrapartum CTGs acquired in singleton term pregnancies, 204 tracings with low signal loss and short time interval to delivery were consecutively selected.

Methods – Tracings were randomly assigned to computer analysis by the Omniview-SisPorto 3.5® system (study group n=104) or to no analysis (control group n=100). Three experienced clinicians evaluated all tracing printouts independently and were asked to predict the newborns’ UApH and 5-minute Apgar scores from them.

Main Outcome Measures – Interobserver agreement, measured by the ICC, and accuracy in prediction of neonatal outcomes with 95CI.

Results – Agreement on prediction of UApH was significantly higher in the study group, with ICC=0.70 [0.61;0.77]95CI, than in the control group, ICC=0.43 [0.21;0.66]95CI, and a trend towards better agreement was also seen in estimation of 5-minute Apgar scores with ICC=0.55 [0.38;0.68]95CI versus ICC=0.43 [0.25;0.57]95CI. Observers predicted UApH values correctly within a 0.10 margin in 70% [0.61;0.79]95CI of cases in the study group versus 46% [0.35;0.56]95CI in the control group. They predicted 5-minute Apgar scores within a margin of one in 81% [0.73;0.88]95CI of cases in the study group and in 70% [0.61;0.79]95CI of cases in the control group.

Conclusions – Prediction of UApH is more reproducible and accurate, when clinicians have access to computerised analysis of CTGs.
INTRODUCTION

Inter and intra-observer variability remains one of the main weaknesses of intrapartum CTG monitoring\(^1\)\(^-\)\(^3\), and cCTG has been proposed as an alternative to overcome this limitation. The Omniview-SisPorto® 3.5 system (Speculum, Lisbon, Portugal) is a program for computer analysis of CTG and ST event signals, developed over the last 21 years at the University of Porto\(^4\)\(^-\)\(^5\). Computerised analysis follows the classical steps of visual CTG analysis: baseline rate estimation, identification of accelerations, decelerations and UC and quantification of short- and long-term variability. The system is extensively described elsewhere\(^5\). Pathological alerts elicited by the Omniview-SisPorto® 3.5 system have been shown to be highly predictive of fetuses born with UA acidemia\(^6\). Despite such promising results, it is unlikely, in the short future, that computerised systems will substitute healthcare professionals in their intrapartum management decisions. Not only would a more extensive evaluation of their validity and effectiveness be required, but also several medical-legal issues need to be adequately addressed.

Management decisions taken by healthcare professionals, based on visual analysis of the intrapartum CTG, have been shown to be poorly reproducible\(^1\)\(^,\)\(^7\), but no studies have looked at this issue, when cCTG is made available. There are also no data on the accuracy of newborn outcome prediction by healthcare professionals, when they are given access to computerised intrapartum CTG.

The aim of this study was to evaluate whether access to cCTG affects clinicians’ reproducibility and accuracy in prediction of newborn UApH and 5-minute Apgar scores. If access to cCTG analysis results in an increased reproducibility and accuracy in prediction of newborn outcome, this would suggest that this methodology could lead to a higher effectiveness. If no such effect should be observed, it would suggest, that improved effectiveness is unlikely, at least as long as management decisions are left in the hands of healthcare professionals.

METHODS

Cases were selected from two pre-existing databases of intrapartum CTG tracings collected for research purposes in a tertiary care university hospital\(^6\)\(^,\)\(^8\). The databases were searched and cases were consecutively selected, if they fulfilled the following criteria: singleton pregnancies, more than 36 completed gestational weeks, fetus in a cephalic presentation, absence of known fetal malformations, active phase of labour, generally accepted indication for internal FHR monitoring (poor signal quality, heavy meconium staining, high-risk pregnancy, etc), a minimum of 60 minutes of tracing duration, signal loss in the last hour under 20%, no complications with the potential to influence fetal oxygenation occurring between tracing end and delivery (difficult vaginal
or abdominal fetal extractions, cord prolapse, maternal hypotension, shoulder dystocia, etc.) and no anaesthetic complications taking place at the time of surgery. Cases were subsequently excluded, if the time interval between tracing end and vaginal delivery exceeded five minutes or if the interval between tracing end and caesarean birth exceeded 20 minutes. In all cases the umbilical cord was doubly clamped immediately after birth and blood was drawn from both artery and vein into previously heparinised syringes. After vestigial air was expelled, blood gas analysis was carried out within 30 minutes after birth. Cases were excluded from UApH analysis, if paired samples were not obtained, if pH values between the two samples differed by less than 0.03 units or if PaCO2 values between the two samples differed by less than 7.5mmHg. Apgar scores were evaluated by the health professional responsible for immediate neonatal support, in the majority of cases this being the attending midwife.

Using computer-generated random numbers, CTG tracings were assigned to receive computer analysis by the Omniview-SisPorto 3.5® system (study group) or no analysis (control group). Computer evaluation of tracings was performed offline, but using a methodology that is similar to real time analysis (i.e. processing starts after the first ten minutes and is subsequently updated every minute, only taking into account signals that were acquired until that point in time). Tracing printout in the study group had the baseline drawn on the FHR graph, accelerations, decelerations, contractions and periods with abnormal long-term and short-term variability highlighted (Figure 1). The last alert elicited by the system was also displayed underneath the tracing. Tracings in the control group only displayed the usual FHR and UC signals. All tracings were printed at a paper speed of 1cm/min and presented independently to three obstetricians with more than five years experience in CTG interpretation. With the information that tracings had been acquired in term pregnancies and that time-intervals to delivery were those previously mentioned, they were asked to estimate the newborns’ UApH (with two decimal places) and 5-minute Apgar scores.
Figure 1: cCTG analysis performed by Omniview-SisPorto® 3.5 system - identification of FHR baseline, accelerations and decelerations with the last real-time alert displayed underneath the tracing.

STATISTICAL METHODS

Interobserver agreement was assessed using the ICC and using limits of agreement (LA), both with 95% CI\(^2\). The accuracy of clinician’s estimations was estimated allowing a maximum 0.10 error for UApH and a one point difference for 5-minute Apgar score. It was evaluated by the percentage of correct estimations and by the agreement between estimated and real values using the LA with three observations per estimation\(^3\). Values of the ICC exceeding 0.75 were interpreted as corresponding to an acceptable agreement. Measures of agreement were calculated using Microsoft® Excel 2003 and SPSS for Windows® version 10.0.7. The t-test and the Man-Whitney test were used for the comparison of the two arms regarding the following parameters: gestational age, birth weight, male births, cord artery pH, cesarean delivery and Apgar score, duration of the tracing, respectively.
RESULTS

The main obstetric characteristics of the study population are displayed in Table 1. Of the 204 tracings selected, 104 were randomised to receive computerised analysis (study group) and 100 to receive no analysis (control group). Five-minute Apgar scores were available in all cases, but valid UAph values were only present in 183 (96 in the study group and 87 in the control group). Clinicians were therefore asked to perform a total of 612 Apgar score estimations and 549 UAph estimations (288 in the study group and 261 in the control group).

The mean value of real UAph was 7.23 with a sd of 0.08. Mean values of UAph predicted by observers A, B and C were 7.22 (sd=0.08), 7.20 (0.06) and 7.18 (0.06), respectively. Real 5-minute Apgar scores ranged from 6 to 10, while values predicted by observers A, B and C ranged from 8-10, 6-10 and 6-10, respectively. Of the three cases of metabolic acidosis (UAph < 7.05 and BDece >12mmol/L) that were included in this study, none went on to develop HIE, and all were assigned to the study group (one was adequately predicted by all observers and the other two were predicted by one of the observers).

Table 1: Randomisation table of the two study groups (visual versus access to eCTG).

<table>
<thead>
<tr>
<th>Analyse</th>
<th>Visual N=100</th>
<th>Computerised N=104</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maen Gestational age in weeks (standard deviation)</td>
<td>39 (1)</td>
<td>39 (1)</td>
</tr>
<tr>
<td>Birth weight in grams mean (standard deviation)</td>
<td>3362 (446)</td>
<td>3282 (427)</td>
</tr>
<tr>
<td>Male births n (%)</td>
<td>50 (50)</td>
<td>46 (44)</td>
</tr>
<tr>
<td>Duration in minutes of the assessed trace median (min-max)</td>
<td>227 (60-770)</td>
<td>213 (64-780)</td>
</tr>
<tr>
<td>Cord artery pH median (min-max) (21 missing values)</td>
<td>7.25 (0.08)</td>
<td>7.22 (0.08)</td>
</tr>
<tr>
<td>5-min Apgar scores median (min-max)</td>
<td>15 (14)</td>
<td>10 (6-10)</td>
</tr>
<tr>
<td>Cesarean delivery (%)</td>
<td>12 (12)</td>
<td>15 (14)</td>
</tr>
</tbody>
</table>

(Legend: N – number of cases; % - percentage of cases).

Table 2 displays the interobserver agreement obtained in prediction of UAph and 5-minute Apgar scores in tracings assigned to the study and to the control groups. Agreement on prediction of UAph, as calculated by the ICC, was significantly higher in the study group. A trend was also seen in this group towards increased interobserver agreement in prediction of 5-minute Apgar scores, but this did not reach statistical significance.
In the study group, observers predicted UApH values correctly, within a 0.10 margin, in 70% of cases [0.61;0.79]95CI, while in the control group this occurred in 46% [0.35;0.56]95CI.

Table 2: Agreement between the three clinicians in prediction of UApH and Apgar scores in tracings with visual CTG analysis and with access to cCTG analysis (95CI in brackets).

<table>
<thead>
<tr>
<th>Limits of agreement</th>
<th>Intraclass correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Limits of agreement</td>
</tr>
<tr>
<td></td>
<td>Visual CTG Analysis</td>
</tr>
<tr>
<td><strong>Umbilical artery blood pH</strong></td>
<td></td>
</tr>
<tr>
<td>Observer A and B</td>
<td>-0.09;0.13</td>
</tr>
<tr>
<td>Observer B and C</td>
<td>-0.15;0.08</td>
</tr>
<tr>
<td>Observer A and C</td>
<td>-0.06;0.19</td>
</tr>
<tr>
<td>Observer A, B and C</td>
<td>-</td>
</tr>
<tr>
<td><strong>5-minute Apgar</strong></td>
<td></td>
</tr>
<tr>
<td>Observer A and B</td>
<td>-1.33;1.49</td>
</tr>
<tr>
<td>Observer B and C</td>
<td>-2.04;1.04</td>
</tr>
<tr>
<td>Observer A and C</td>
<td>-0.98;2.14</td>
</tr>
<tr>
<td>Observer A, B and C</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 2 displays the individual differences between predicted and real pH values, together with the obtained LA values in the study (-0.16;0.11) and in the control group (-0.21;0.14). For the 5-minute Apgar score, correct predictions, within a margin of one, were obtained in 81% [0.73;0.88]95CI of cases in the study group versus 70% of cases [0.61;0.79]95CI in the control group.
Figure 2: Observer-outcome agreement in prediction of birth UApH using Bland and Altman LA. Three observations per tracing with 95CI in FHR tracings with and without cCTG analysis.

Table 3 displays the agreement between observers and real UApH and also between all three observers and the real UApH, using the ICC as the statistical measure.

Table 3: Agreement between the three clinicians in prediction of UApH in tracings with visual CTG analysis and with access to cCTG analysis (95CI in brackets).

<table>
<thead>
<tr>
<th></th>
<th>Visual CTG analysis</th>
<th>cCTG analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Umbilical artery pH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A and real</td>
<td>0.36 [0.16;0.53]</td>
<td>0.54 [0.38;0.67]</td>
</tr>
<tr>
<td>B and real</td>
<td>0.31 [0.10;0.49]</td>
<td>0.54 [0.35;0.68]</td>
</tr>
<tr>
<td>C and real</td>
<td>0.12 [-0.05;0.30]</td>
<td>0.33 [0.13;0.50]</td>
</tr>
<tr>
<td>All observers and real</td>
<td>0.29 [0.08;0.47]</td>
<td>0.52 [0.34;0.66]</td>
</tr>
</tbody>
</table>
DISCUSSION

This study demonstrates that clinicians agree more with each other on prediction of UAB pH and are more accurate in this prediction when they have access to computerised analysis of CTG tracings. This suggests that clinicians are consciously or unconsciously influenced by the results of cCTG analysis and perhaps it leads to a more homogeneous and more correct tracing interpretation. However, it is possible that the degree of influence will depend on their previous experience and on their personal confidence with the system.

Few studies have addressed the accuracy of healthcare professionals in predicting UApH and Apgar scores based on the intrapartum CTG. Chauhan et al\(^4\) evaluated the accuracy of five clinicians in estimating UApH < 7.00, BE ≥ 12 mmol/l and Apgar scores ≤ 3 at 5 minutes by visual analysis of CTG tracings. One-hundred intrapartum non-reassuring FHR tracings were reviewed, from one hour prior to the appearance of CTG abnormalities and, if applicable, the hour before delivery. Large discrepancies were found in prediction of neonatal outcome variables. Overall Spearman correlation coefficients for prediction of low Apgar score, low pH and abnormal BE ranged from 0.11 to 0.19, demonstrating no positive association between predicted and real outcomes. This could have been due to the poor interobserver agreement found in visual classification of CTG tracings (wk ranging from -0.12 to 0.15). This study concluded that visual analysis of intrapartum CTGs is not a useful diagnostic test for the identification of fetuses born with low Apgar score or abnormal acid-base state (likelihood ratio 1-2).

Nielsen et al\(^5\) compared the prediction of fetal outcome obtained by a computer system with that of four experienced obstetricians performing visual analysis of CTGs. The final 30 minutes of 50 intrapartum tracings were evaluated. A dichotomised classification of fetal outcome as normal or compromised was used, the later defined as a one-minute Apgar score <7, UApH <7.15, BE <-10mmol/l or the need for primary resuscitation. The computer system obtained an accuracy of 86%, which was significantly higher than that of obstetricians (50-66%).

One may consider that a correct prediction of UApH values, within a 0.10 margin, in 70% of tracings is not impressive, particularly given the sd in this value of 0.08. The distribution of UA values in the general population may lead to similar results if an average value is always predicted. Nevertheless, the accuracy of prediction was significantly higher in the study group and the only possible explanation for this is the access to cCTG results. Prediction of Apgar scores from CTG tracings has a more limited value, as they are known to be affected by several other factors than oxygenation and they are subject to high interobserver variability. It was interesting to find non-significant trends towards a higher reproducibility and accuracy of 5-minute
Apgar score prediction in the group that had access to cCTG. However it is not known whether a larger sample size would lead to a significant result.

When evaluating clinicians’ accuracy in prediction of neonatal outcome parameters the decision was made to assess precision within a margin of error, rather than grouping cases into dichotomous classes and evaluating the sensibility and specificity. The wide variation in normal UAph values is well known, and attention has traditionally focused on cases with pH <7.05 or pH <7.00, particularly when associated with BD exceeding 12mmol/l (metabolic acidosis). However, this approach requires a much larger sample size and/or an artificial selection of poor outcome cases. Evaluating the degree of fetal acidemia, rather than identification of dichotomous classes is also a clinically useful objective, as this is frequently employed when deciding the timing of a clinical intervention, which should ideally be performed before the onset of metabolic acidosis.

It is well known that fetal oxygenation can deteriorate rapidly in unstable intrapartum situations. In this study great care was taken to reduce periods of signal loss and the CTG-delivery time interval to a minimum, so that the studied CTG pattern reflected as much as possible the fetal oxygenation at the time of delivery, and hence the five minute interval to vaginal birth and 20 minute interval to caesarean section. In spite of this, it is acknowledged that particularly the latter period could have introduced some uncertainty into the results.

**CONCLUSION**

We have shown that access to cCTG analysis improves prediction of UAph. This has the potential to improve clinicians’ management decisions based of the intrapartum CTG. Whether this will reduce the incidence of adverse neonatal outcomes needs to be adequately evaluated in a RCT.

**REFERENCES**


Prediction of neonatal acidemia by computer analysis of fetal heart rate and ST event signals.

Antónia Costa¹², Diogo Ayres-de-Campos¹², Fernanda Costa³, Cristina Santos⁴⁵, João Bernardes¹²

¹ Department of Obstetrics and Gynaecology, Faculty of Medicine, Porto University, São João Hospital, Portugal
² Institute of Biomedical Engineering, Porto
³ Department of Obstetrics and Gynaecology, São João Hospital, Portugal
⁴ Department of Biostatistics and Medical Informatics, Faculty of Medicine, Porto University
⁵ Centre for Research in Health Technologies and Information Systems, Faculty of Medicine, Porto University
**ABSTRACT**

Objective: Evaluate the accuracy of computer analysis of FHR and ST event signals in prediction of neonatal acidemia.

Study design: 148 FHR tracings were evaluated to identify “red alerts” provided by the system, based on automated analysis of FHR and ST event signals and compared with the occurrence of UA acidemia (pH ≤ 7.05).

Results: Presence of “red alerts” obtained sensitivity of 1.00 [0.56;1] 95CI, specificity of 0.94 [0.89;0.97] 95CI, PPV of 0.47 [0.22;0.72] 95CI, NPV of 1 [0.96;1.00] 95CI, positive likelihood ratio (PLR) of 17.6 [9.0;34.5] 95CI and negative likelihood ratio (NLR) of 0. When limiting analysis to “red alerts”, that did not include ST data, sensitivity was 0.57 [0.20;0.88] 95CI, specificity 0.97 [0.92;0.99] 95CI, PPV 0.50 [0.17;0.82] 95CI, NPV 0.98 [0.93;0.99] 95CI, PLR 20.14 [6.3;64.2] 95CI and NLR 0.44 [0.19;1.04] 95CI.

Conclusions: Computer analysis of FHR and ST event signals provide higher accuracy in predicting neonatal acidemia.
INTRODUCTION

FHR monitoring remains widely used as a method for detecting changes in fetal oxygenation that can occur during labour. Yet deaths and long-term disablement from intrapartum hypoxia remain an important cause of suffering for parents and families, even in industrialised countries\(^1\). Confidential inquiries in the United Kingdom have highlighted that as much as 50% of these deaths could have been avoided, as they were caused by non-recognition of abnormal FHR patterns, poor communication between staff or delay in taking appropriate action\(^2\). Analysis of FHR tracings is almost universally performed visually by health professionals, and this has been shown to be subject to wide intra- and interobserver variation, both in what concerns overall interpretation\(^3,4\) and identification of individual tracing events\(^5,6\).

More recently, monitoring of fetal electrocardiographic ST waveform signals (STAN\(^\text{®}\), Neoventa, Gothenburg, Sweden) has arisen as an adjunct to intrapartum FHR monitoring and has been shown to reduce unnecessary interventions caused by a high incidence of false positive FHR patterns. A systematic review of the first three conducted trials comparing joint FHR+ST monitoring with isolated FHR monitoring showed, that the former significantly decreases the rates of fetal blood sampling, neonatal encephalopathy, operative delivery and the incidence of UA metabolic acidosis\(^7\). The technique is gaining widespread acceptance and has been licensed by the Food and Drug Administration. However, adverse neonatal outcomes continue to occur with routine use of the STAN\(^\text{®}\) technology, mainly because of human errors, such as poor FHR tracing interpretation, delay in taking appropriate action, failure to follow clinical guidelines\(^8\) and, in rare cases, non-occurrence or very late occurrence of ST events\(^9\). A recent consensus statement from Europe identified visual interpretation of FHR tracings as the technology’s main weakness\(^10\).

Computerized analysis of FHR tracings has been developed over the last three decades, as a way to overcome the poor reproducibility of visual analysis. Most systems have focused in analysis of antepartum tracings\(^11,12\) with several limitations being reported when it is applied in the intrapartum\(^12,13\). Intrapartum tracings pose additional difficulties for computerized analysis, related to increased tracing length and hence additional computer memory requirements, frequent occurrence of signal loss and artefacts, and greater signal instability resulting in more complicated baseline estimation.

Omniview-SisPorto\(^\text{®}\) 3.5 (Speculum\(^\text{CE}\), Lisbon, Portugal) is the latest version\(^14,15\) of a system developed for analysis of both antepartum and intrapartum signals\(^16\). It incorporates the concepts of a FHR central monitoring station with multiple viewing posts, and a system for computerised analysis that exhibits real-time alerts for combined FHR+ST event abnormalities. This technology aims to reduce human error in tracing interpretation and possible delay in taking appropriate action, by providing
health professionals with visual and sound alerts to FHR changes. The system’s analysis of abnormal FHR variability can also be an adjunct to the detection of fetal hypoxia, particularly in cases that do not display ST alerts.

The aim of this study was to evaluate the predictive capacity of the Omniview-SisPorto® 3.5 system, when analysing the last hour of intrapartum STAN® tracings, to identify cases that went on to display UA acidemia at birth (i.e. UApH ≤ 7.05).

**MATERIAL AND METHODS**

This prospective observational study was conducted in a tertiary care university hospital between May 2005 and September 2007. Local ethical committee approval for the study was provided and written informed consent for enrolment was obtained from all subjects.

Consecutive cases were enrolled, if they fulfilled the following inclusion criteria: singleton pregnancy, more than 36 completed gestational weeks, fetus in cephalic presentation, absence of known fetal malformations, active phase of labour and a generally accepted indication for internal FHR monitoring (poor signal quality, heavy meconium staining and high-risk pregnancy). All patients were continuously monitored until delivery with FHR+ST analysis using either a STAN® 21 or a STAN® 31 monitor.

Enrolled cases were subsequently excluded, if one the following situations occurred: tracing lasting less than 60 minutes, signal loss in the last hour exceeding 15%, complications with the potential to influence fetal oxygenation between tracing-end and delivery (such as difficult vaginal or abdominal fetal extractions, cord prolapse, maternal hypotension and shoulder dystocia), anesthetic complications taking place at the time of surgery or umbilical cord blood values that were considered inadequate. For practical reasons related to the time needed for application of a ventouse or for the preparation of a cesarean section, cases in which the interval between tracing-end and vaginal delivery exceeded 5 minutes or until cesarean birth exceeded 20 minutes were excluded. In all cases the umbilical cord was doubly clamped immediately after birth, and blood was aspirated from both artery and vein into previously heparinised syringes. After vestigial air was expelled, blood gas analysis was carried out within 30 minutes after birth. Cases were excluded from the analysis, if paired samples were not obtained, if pH values between the two samples differed by less than 0.03 units or if PaCO2 values between the two samples differed by less than 7.5 mmHg. Apgar scores were evaluated at 1 and 5 minutes by the health professional responsible for immediate neonatal support, in the majority of cases this being a specialised labour and delivery nurse. In all cases with UApH ≤ 7.05, 5-minute Apgar score < 5 or NICU admission, the newborn’s hospital records were reviewed by one of the authors (AC) for the diagnosis of HIE, as established by the attending neonatologist. This was defined as the
appearance of changes in muscle tone, feeding, state of conscience or seizures occurring in the first 48 hours of life associated with laboratory evidence of peripartum acidemia.

Computer analysis of CTG+ST tracings by the Omniview-SisPorto 3.5® system was performed offline for this study, but using a similar methodology to the one that is employed, when tracings are being acquired in real time (i.e. only signals acquired until the present moment are available to elicit alarms). This was accomplished using a program feature, by which signals are processed from the start of the file, one minute at a time, only taking into account signals acquired until then, and a log of alarms elicited on a minute-by-minute basis is recorded. Thus, no alerts that would be elicited during real tracing acquisition are lost or changed.

The last hour of the tracing was reviewed by one of the authors (A.C), and the “red alerts” provided by the system during that period were compared with the occurrence of UA acidemia at birth, defined as a pH value ≤ 7.05. “Red alerts” are elicited by the system, when it detects FHR+ST changes that are considered to be strongly associated with fetal hypoxemia. These include “reduced long-term variability”, “repetitive decelerations and reduced long-term variability”, “tachycardia, decelerations and ST event”, “decelerations and ST event”, “repetitive decelerations and ST event”, “very repetitive decelerations and ST event”, “prolonged decelerations and ST event”, “reduced short-term variability”, “repetitive decelerations and reduced short-term variability”, “very prolonged deceleration” and “prolonged deceleration with reduced variability”. Changes in FHR+ST signals that are considered less serious, will elicit an “orange” or “yellow” alert. A more detailed description of these alerts is available elsewhere.

STATISTICAL METHODS

Sensitivity, specificity, positive and negative predictive values, positive and negative likelihood ratios were calculated with 95% confidence intervals (95% CI), using Microsoft® Excel 2003 and SPSS for Windows® version 10.0.7.

RESULTS

A total of 193 consecutive patients were enrolled, and 45 were subsequently excluded: 19 due to disconnection of the scalp electrode conditioning a large tracing-to-delivery interval and/or signal loss, 21 due to inadequate umbilical blood gases values and five due to insufficient tracing duration. This resulted in a total of 148 cases.
available for analysis. Nineteen had maternal risk factors (eight hypertensive diseases, eight diabetes, one anemia and two thrombophilias), four had fetal risk factors (2 hydramnios, 1 oligohydramnios, 1 fetal growth restriction and 9 heavy meconium staining), 53 cases had an induced labour and 16 cases had prolonged rupture of membranes (>16 hours). The remaining cases were monitored internally due to poor FHR signal quality with external acquisition methods.

The main characteristics of the study population are displayed in table 1. Mean newborn birth weight was 3231g with a sd of 387g. Median gestational age at delivery was 39 weeks. Median duration of the tracings was 171 minutes with a minimum of 60 minutes and a maximum of 780 minutes.

Table 1: Description of gestational age, type of delivery, Apgar scores and newborn gender of the 148 study cases.

<table>
<thead>
<tr>
<th>Gestational age in weeks</th>
<th>Number of cases</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>2</td>
<td>(1%)</td>
</tr>
<tr>
<td>37</td>
<td>10</td>
<td>(7%)</td>
</tr>
<tr>
<td>38</td>
<td>39</td>
<td>(26%)</td>
</tr>
<tr>
<td>39</td>
<td>44</td>
<td>(30%)</td>
</tr>
<tr>
<td>40</td>
<td>39</td>
<td>(26%)</td>
</tr>
<tr>
<td>41</td>
<td>14</td>
<td>(10%)</td>
</tr>
<tr>
<td>Type of delivery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vaginal</td>
<td>73</td>
<td>(49%)</td>
</tr>
<tr>
<td>instrumental</td>
<td>43</td>
<td>(29%)</td>
</tr>
<tr>
<td>cesarean</td>
<td>32</td>
<td>(22%)</td>
</tr>
<tr>
<td>1-minute Apgar score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;4</td>
<td>1</td>
<td>(1%)</td>
</tr>
<tr>
<td>4-6</td>
<td>8</td>
<td>(5%)</td>
</tr>
<tr>
<td>&gt;6</td>
<td>139</td>
<td>(94%)</td>
</tr>
<tr>
<td>5-minute Apgar score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;4</td>
<td>0</td>
<td>(0%)</td>
</tr>
<tr>
<td>4-6</td>
<td>1</td>
<td>(1%)</td>
</tr>
<tr>
<td>&gt;6</td>
<td>147</td>
<td>(99%)</td>
</tr>
<tr>
<td>Newborn gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>75</td>
<td>(51%)</td>
</tr>
<tr>
<td>Male</td>
<td>73</td>
<td>(49%)</td>
</tr>
</tbody>
</table>

There were seven cases of UApH \( \leq 7.05 \), and these are displayed in detail in table 2. No cases of neonatal HIE occurred in this series.
Table 2: Cases displaying “red alerts” (including FHR+ST signal features) and isolated CTG alert in the last hour of tracing with their neonatal outcome parameters.

<table>
<thead>
<tr>
<th>FHR alert</th>
<th>ST event</th>
<th>Alert</th>
<th>CTG alert</th>
<th>Delivery</th>
<th>Lag-time</th>
<th>UA pH</th>
<th>UA BDecf</th>
<th>UV pH</th>
<th>UV BDecf</th>
<th>BW</th>
<th>Apgar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep D</td>
<td>T/QRS rise</td>
<td>Yes</td>
<td>Orange</td>
<td>Inst</td>
<td>5</td>
<td>6.90</td>
<td>12</td>
<td>7.02</td>
<td>12</td>
<td>3160</td>
<td>7/8</td>
</tr>
<tr>
<td>Prol D + ▼ V</td>
<td>T/QRS rise</td>
<td>Yes</td>
<td>Red</td>
<td>CS</td>
<td>4</td>
<td>6.94</td>
<td>14</td>
<td>7.04</td>
<td>14</td>
<td>2560</td>
<td>7/9</td>
</tr>
<tr>
<td>Rep D</td>
<td>BP2/3 ST</td>
<td>Yes</td>
<td>Orange</td>
<td>CS</td>
<td>10</td>
<td>6.97</td>
<td>10</td>
<td>7.05</td>
<td>8</td>
<td>2850</td>
<td>3/8</td>
</tr>
<tr>
<td>Prol D + ▼ V</td>
<td>ST signal loss</td>
<td>Yes</td>
<td>Red</td>
<td>Inst</td>
<td>4</td>
<td>6.98</td>
<td>8</td>
<td>7.01</td>
<td>11</td>
<td>3250</td>
<td>7/9</td>
</tr>
<tr>
<td>Rep D</td>
<td>T/QRS rise</td>
<td>Yes</td>
<td>Orange</td>
<td>Vag</td>
<td>2</td>
<td>7.04</td>
<td>10</td>
<td>7.07</td>
<td>11</td>
<td>3270</td>
<td>7/9</td>
</tr>
<tr>
<td>Prol D + ▼ V</td>
<td>ST signal loss</td>
<td>Yes</td>
<td>Red</td>
<td>Vag</td>
<td>0</td>
<td>7.05</td>
<td>10</td>
<td>7.17</td>
<td>11</td>
<td>2910</td>
<td>7/9</td>
</tr>
<tr>
<td>Prol D + ▼ V</td>
<td>T/QRS rise</td>
<td>Yes</td>
<td>Red</td>
<td>Vag</td>
<td>0</td>
<td>7.06</td>
<td>9</td>
<td>7.11</td>
<td>8</td>
<td>2850</td>
<td>6/10</td>
</tr>
<tr>
<td>Rep D</td>
<td>BP2/3 ST</td>
<td>Yes</td>
<td>Yellow</td>
<td>CS</td>
<td>8</td>
<td>7.09</td>
<td>8</td>
<td>7.15</td>
<td>9</td>
<td>4430</td>
<td>7/9</td>
</tr>
<tr>
<td>Rep D</td>
<td>BP2/3 ST</td>
<td>Yes</td>
<td>Red</td>
<td>CS</td>
<td>12</td>
<td>7.11</td>
<td>7</td>
<td>7.15</td>
<td>7</td>
<td>3500</td>
<td>9/10</td>
</tr>
<tr>
<td>Rep D</td>
<td>T/QRS rise</td>
<td>Yes</td>
<td>Yellow</td>
<td>Vag</td>
<td>0</td>
<td>7.13</td>
<td>9</td>
<td>7.21</td>
<td>9</td>
<td>2586</td>
<td>9/10</td>
</tr>
<tr>
<td>Rep D</td>
<td>T/QRS rise</td>
<td>Yes</td>
<td>Yellow</td>
<td>CS</td>
<td>17</td>
<td>7.13</td>
<td>7</td>
<td>7.25</td>
<td>11</td>
<td>2090</td>
<td>9/10</td>
</tr>
<tr>
<td>Rep D</td>
<td>T/QRS rise</td>
<td>Yes</td>
<td>Yellow</td>
<td>Vag</td>
<td>0</td>
<td>7.16</td>
<td>6</td>
<td>7.27</td>
<td>2</td>
<td>3590</td>
<td>9/10</td>
</tr>
<tr>
<td>Rep D</td>
<td>T/QRS rise</td>
<td>Yes</td>
<td>Yellow</td>
<td>Vag</td>
<td>0</td>
<td>7.28</td>
<td>7</td>
<td>7.40</td>
<td>7</td>
<td>3146</td>
<td>9/10</td>
</tr>
<tr>
<td>Rep D</td>
<td>BP2/3 ST</td>
<td>Yes</td>
<td>Yellow</td>
<td>Vag</td>
<td>0</td>
<td>7.28</td>
<td>1</td>
<td>7.32</td>
<td>2</td>
<td>3460</td>
<td>9/10</td>
</tr>
<tr>
<td>LT V</td>
<td>No</td>
<td>Yes</td>
<td>Red</td>
<td>CS</td>
<td>20</td>
<td>7.32</td>
<td>1</td>
<td>7.39</td>
<td>2</td>
<td>3830</td>
<td>9/10</td>
</tr>
<tr>
<td>ST V</td>
<td>No</td>
<td>Yes</td>
<td>Red</td>
<td>Vag</td>
<td>0</td>
<td>7.33</td>
<td>4</td>
<td>7.40</td>
<td>5</td>
<td>3390</td>
<td>8/9</td>
</tr>
</tbody>
</table>

(Legend: Alert – system red alert; CTG alert – system CTG alert without ST integration; Delivery – type of delivery; Vag – normal vaginal delivery; Inst -instrumental vaginal delivery; CS - cesarean section; Lag-time – time interval in minutes between end of FHR monitoring and delivery; UA – umbilical artery; UV – umbilical vein; BDecf - base deficit in the extracellular fluid in mmol/l; BW – birth weight in grams; Apgar – 1-minute/5-minute Apgar score; Rep D – repetitive deceleration; Prol D + ▼ V – prolonged deceleration with reduced variability; VP D – Very prolonged deceleration; Rep D + ▼ STV – repetitive deceleration with reduced short term variability; ▼ LTV – reduced long term variability; ▼ STV – reduced short term variability; BP2/3 ST – biphasic event type 2 or 3).

Table 3 compares the appearance of “red alerts” in the last 60 minutes with the occurrence of umbilical artery acidemia.
Table 3: Number of cases displaying the system’s “red alerts” and isolated CTG “red alerts” with corresponding UApH values above or below the selected cut-off.

<table>
<thead>
<tr>
<th></th>
<th>pH</th>
<th>≤ 7.05</th>
<th>&gt;7.05</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>FHR+ST “red alerts”</td>
<td>Yes</td>
<td>7</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>0</td>
<td>133</td>
<td>133</td>
</tr>
<tr>
<td>Isolated FHR “red alerts”</td>
<td>Yes</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>3</td>
<td>137</td>
<td>140</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>7</td>
<td>141</td>
<td>148</td>
</tr>
</tbody>
</table>

In the 148 studied cases the occurrence of “red alerts” in the last hour of the tracing obtained a sensitivity of 1 [0.56;1.00]$_{95}$CI, a specificity of 0.94 [0.89;0.97]$_{95}$CI, a PPV of 0.47 [0.22;0.72]$_{95}$CI, a NPV of 1 [0.96;1.00]$_{95}$CI, a positive likelihood ratio of 17.6 [9.0;34.5]$_{95}$CI and a negative likelihood ratio of 0 in prediction of newborn acidemia. When evaluating “red alerts”, elicited exclusively by FHR analysis and not incorporating ST data, the obtained sensitivity was 0.57 [0.20;0.88]$_{95}$CI, the specificity 0.97 [0.92;0.99]$_{95}$CI, the PPV 0.50 [0.17;0.82]$_{95}$CI, the NPV 0.98 [0.93;0.99]$_{95}$CI, the positive likelihood ratio 20.14 [6.3;64.2]$_{95}$CI and the negative likelihood ratio 0.44 [0.19;1.04]$_{95}$CI.

COMMENT

This is the first study to evaluate the accuracy of intrapartum computerized analysis of FHR+ST signals. Confidence intervals for sensitivity and positive predictive value are still wide, due to the occurrence of a small number of cases with newborn acidemia. Further research is needed in order to confirm the high sensitivity found in this study. On the other hand, the small confidence intervals, surrounding the high specificity and negative predictive values, provide promising results for this technology.

Only a small number of studies have evaluated the accuracy of computer analysis of intrapartum FHR signals in prediction of neonatal outcome, and all were carried out more than 13 years ago. Keith et al.\textsuperscript{25} compared the results of a computer system with 17 experts in the interpretation of 50 intrapartum FHR tracings. The system showed a fair agreement with experts and did not recommend unnecessary interventions in babies born in good condition (cord artery pH > 7.15, vein pH > 7.20, 5- minute Apgar ≥ 9 and no resuscitation). It identified as many birth-asphyxiated cases (cord arterial pH < 7.05 and BD$_{ecf}$ ≥ 12mmol/l and Apgar score at 5 min ≤ 7 with neonatal morbidity) as...
the majority of experts. However, in one of these cases the system the majority of experts failed to recommend intervention. No further evaluations of this technology have, to our knowledge, been published. Chung et al evaluated the ability of another computer system for FHR tracing analysis to predict newborn acidemia, defined as an UApH under 7.15 (n=8). In a set of 73 intrapartum tracings, a sensitivity of 88% and a specificity of 75% were reported. For prediction of an UA BD>8 mmol/l (n=17) the sensitivity was 76% and the specificity 82%. Nielsen et al evaluated the accuracy of a third computer system by analysing the last 30 minutes of 50 intrapartum FHR tracings. Defining poor fetal outcome as a 1-minute Apgar score below 7, UApH below 7.15, standard BE below -10 meq/l and the need for primary resuscitation (n=16), the overall accuracy was 72%, while four obstetricians, who were assigned the same task, reached an accuracy of 64%.

Recent evidence using non-computerised assessment has shown, that the addition of ST segment analysis can enhance the accuracy of fetal monitoring, and our results also suggest this hypothesis. Using the system’s “red alerts”, that incorporate both FHR and ST event features, all seven acidic fetuses were identified, and there were eight false positive results (Table 2). Six of these cases had an UApH under 7.20, suggesting that a milder degree of hypoxemia was occurring. When ST event data were excluded, only four out of the seven acidic cases were identified, and there were an additional four false positives. Thus, three acidic fetuses would not have displayed “red alerts” if isolated FHR monitoring had been performed. The latter has the tradition of providing a close to 100% detection rate, albeit with a limited positive predictive value. It can be seen from table 2, that all cases would have been detected if the “orange alerts”, the system’s second most ominous alert, had also been included. The appearance of this alert should prompt further investigation, including fetal blood sampling (FBS) or STAN® monitoring. However, inclusion of “orange alerts” for evaluation of the system’s accuracy would also cause a sharp increase in the number of false positives. False positives are the main drawback of visual FHR analysis and also occur with the STAN® technology. They may be caused by fetal adaptations to the challenges of labour, rather than by noxious stimuli. The system appears to have reduced, but not eliminated these cases, suggesting that further refinement of some alerts will be required.

Analysis of alerts was only carried out in the last 60 minutes of the tracing, as this was thought the bear the highest relationship with immediate newborn outcome. Additional alerts were present in a small number of cases before that period, both in cases with and without umbilical artery acidemia. However, these were not included, as they could have elicited an intervention, that led to the reversion of what was only a temporary acidemia.
The rationale for choosing UA acidemia with a cut-off value of 7.05 as the main outcome parameter for this study was a compromise between clinical significance and overall incidence. Metabolic acidosis, defined as an UApH ≤ 7.05 and a BD ecf exceeding 12 mmol/l, would have been preferable, from the clinical point of view, but its low incidence in intensely monitored populations (only two cases in the study sample), leads to the need for enrolment of a much larger number of cases. The isolated use of the pH value has the drawback of including cases of severe respiratory acidosis, which are much more rapidly reversible and less meaningful in terms of hypoxic risk for the fetus. Apgar score was not included as a measure of outcome, because of its subjective components and poor correlation with short- and long-term fetal outcomes28-30.

The high number of excluded cases was a consequence of the strict criteria required for analysis, and these were considered necessary in order to evaluate the true potential of the diagnostic test without the various biases that affect its performance in everyday clinical practice, such as inadequate tracing length, signal loss and long interval to birth. These are likely to reduce its overall accuracy and, therefore, further studies are needed to evaluate this characteristic in routine clinical practice.

In conclusion, computer analysis assures a robust approach to the issue of accuracy of FHR monitoring in prediction of neonatal outcome. The present study demonstrates that a high accuracy may be obtained with computer analysis of FHR+ST signals in prediction of term fetuses that are born with UA acidemia. Although this was the largest sample size used to address this issue, confidence intervals for specificity and PPV are still wide, and further studies are needed to confirm this finding.

The system’s alerts, which appear together with a description of the criteria that were identified to evoke them, are expected to capture the health professional’s attention and to provoke a re-evaluation of the case under a new light, but they do not obviously guarantee a satisfactory outcome. The effect of introducing such a system into routine clinical practice requires the conduction of an adequately sized RCT, which is currently being prepared.

REFERENCES


A randomised clinical trial of intrapartum fetal monitoring with computer analysis and alerts versus previously available monitoring – Study protocol.

Diogo Ayres-de-Campos\textsuperscript{1,2}, Austin Ugwumadu\textsuperscript{3}, Philip Banfield\textsuperscript{4}, Pauline Lynch\textsuperscript{5}, Pina Amin\textsuperscript{6}, David Horwell\textsuperscript{7}, Antonia Costa\textsuperscript{1,2}, Cristina Santos\textsuperscript{8,9}, João Bernardes\textsuperscript{1,2}, Karl Rosen\textsuperscript{10}

\textsuperscript{1} Department of Obstetrics and Gynaecology, Faculty of Medicine, Porto University, São João Hospital, Portugal
\textsuperscript{2} Institute of Biomedical Engineering, Porto
\textsuperscript{3} Department of Obstetrics and Gynaecology, St. George’s Hospital, University of London, UK
\textsuperscript{4} Department of Obstetrics and Gynaecology, Glan Clwyd Hospital, Wales, UK
\textsuperscript{5} Department of Obstetrics and Gynaecology, Ninewells Hospital, Dundee, Scotland, UK
\textsuperscript{6} Department of Obstetrics and Gynaecology, University Hospital of Wales, Wales, UK
\textsuperscript{7} Department of Obstetrics and Gynaecology, Luton and Dunstable Hospital, UK
\textsuperscript{8} Department of Biostatistics and Medical Informatics, Faculty of Medicine, Porto University
\textsuperscript{9} Centre for Research in Health Technologies and Information Systems, Faculty of Medicine, Porto University
\textsuperscript{10} Department of Engineering, University of Borås, Sweden
ABSTRACT

Background: Intrapartum fetal hypoxia remains an important cause of death and permanent handicap, and, in a significant proportion of cases, there is evidence of suboptimal care related to fetal surveillance. CTG monitoring remains the basis of intrapartum surveillance, but its interpretation by healthcare professionals lacks reproducibility and the technology has not been shown to improve clinically important outcomes. The addition of fECG analysis has increased the potential to avoid adverse outcomes, but CTG interpretation remains its main weakness. A program for computerised analysis of intrapartum fetal signals, incorporating real-time alerts for healthcare professionals, has recently been developed. There is a need to determine whether this technology can result in better perinatal outcomes.

Methods/Design: This is a multicentre RCT. Inclusion criteria are: women aged ≥ 16 years, able to provide written informed consent, singleton pregnancies ≥ 36 weeks, cephalic presentation, no known major fetal malformations, in labour, but excluding active second stage, planned for continuous CTG monitoring and no known contra-indication for vaginal delivery. Eligible women will be randomised using a computer-generated randomization sequence to one of the two arms: continuous computer analysis of fetal monitoring signals with real-time alerts (intervention arm) or continuous CTG monitoring as previously performed (control arm). Electrocardiographic monitoring and fetal scalp blood sampling (FBS) will be available in both arms. The primary outcome measure is the incidence of fetal metabolic acidosis (UApH <7.05, BD_{ref} >12 mmol/l). Secondary outcome measures are: caesarean section and instrumental vaginal delivery rates, use of fetal blood sampling, 5-minute Apgar score <7, NICU unit admission, moderate and severe neonatal encephalopathy with a marker of hypoxia, perinatal death, rate of internal monitoring, tracing quality and signal loss. Analysis will follow an intention to treat principle. Incidences of primary and secondary outcomes will be compared between groups. Assuming a reduction in metabolic acidosis from 2.8% to 1.8%, using a two-sided test with $\alpha=0.05$, power=0.80, and 10% loss to follow-up, 8133 women need to be randomised.

Discussion: This study will provide evidence of the impact of intrapartum monitoring with computer analysis and real-time alerts on the incidence of adverse perinatal outcomes, intrapartum interventions and signal quality (Current controlled trials ISRCTN42314164).
BACKGROUND

Intrapartum complications accounted for 512 perinatal deaths in the UK in 2004 and remain an important cause of long-term neurological morbidity and suffering for families in industrialised countries. In more than half of such cases there is evidence of suboptimal care, where different management would reasonably have been expected to have made a difference to the outcome. Problems related to fetal surveillance are the most commonly reported in these cases, and CTG interpretation is the basis for the most frequent criticism.

The aim of intrapartum monitoring is to identify fetuses at risk of death or long-term injury caused by decreased oxygen supply during labour. The majority of cases of death and long-term disability are caused by situations other than poor oxygenation during labour. So to establish this diagnosis, it is necessary to document the occurrence of relevant changes in umbilical blood gas values after birth. UA metabolic acidosis has been associated with an increased risk of neurological injury and is commonly used as a proxy measure for adverse clinical outcome in this setting.

CTG monitoring remains the basis of intrapartum fetal surveillance in high-risk cases and is applied on a wide scale in industrialised countries, but its interpretation by health professionals has a well documented poor reproducibility, and the technology has not been shown to improve the most important clinical outcomes, but rather to increase operative delivery rates. FBS can be used in addition to CTG, but it is invasive and time consuming and only provides time-limited information, all of which have limited its application on a wide scale.

The addition of ECG ST waveform analysis to conventional CTG (STAN®, Neoventa, Gothenburg, Sweden) has been shown to increase the identification of fetuses with metabolic acidosis. A systematic review of the first three trials comparing CTG+ST monitoring with conventional CTG showed that the former significantly decreases the rates of FBS, neonatal encephalopathy and operative delivery and is associated with a borderline reduction in the incidence of UA metabolic acidosis. It has recently been documented, that adverse neonatal outcomes continue to occur with routine use of the STAN® technology, because of human errors, such as poor CTG interpretation, delay in taking appropriate action or failure to follow clinical guidelines, as well as non-occurrence or very late occurrence of ST events. There is now a consensus among STAN® users that visual interpretation of the CTG remains the main weakness of the technology.

The Omniview-SisPorto® 3.5 program (Speculum, Lisbon, Portugal) provides computer analysis of both CTG and ST signals, incorporating the concept of centralised viewing of tracings on multiple stations and real-time alerts for healthcare professionals. The system has been shown to provide analysis of CTG events, that is
in good agreement with a consensus of experts\textsuperscript{15} and the program’s alerts have been shown to be highly predictive of fetuses born with severe acidemia\textsuperscript{16}, so it has the potential to overcome some of the weaknesses associated with human interpretation of the CTG. There is now a need to determine whether the use of this technology will result in improved perinatal outcomes.

**AIMS**

The primary aim of the study is to determine, whether computer analysis of intrapartum fetal monitoring signals with real-time alerts (Omniview-SisPorto\textsuperscript{®} 3.5) will reduce the rate of UA metabolic acidosis compared to continuous electronic fetal monitoring as previously performed. Secondary aims are to quantify other measures of perinatal outcome, intervention rates and signal quality measures in both arms of the study.

The rationale for the main hypothesis of the study is that real-time alerts are expected to prompt healthcare professionals to identify and act on changes that would otherwise remain unnoticed. The technology may also reduced human errors associated with inappropriate CTG interpretation, including lack of identification of cases showing reduced FHR variability\textsuperscript{12}. Observational data suggest that the validity of computerised intrapartum fetal monitoring provides additional advantages\textsuperscript{16}.

Written informed consent for enrolment will be requested from all participants. The trial is registered at Current Controlled Trials with the number ISRCTN42314164. The research protocol has been submitted for evaluation by the United Kingdom Integrated Research Application System.
STUDY DESIGN AND SETTING

This is a pragmatic multicentre RCT to be carried out in five United Kingdom hospitals, including three tertiary teaching units and two district general hospitals, all with high-risk women in labour.

POPULATION AND METHODS

PARTICIPANTS/ELIGIBILITY CRITERIA

Women will be eligible for participation if they fulfill the following criteria:

- singleton pregnancy with cephalic presentation
- gestation ≥ 36 completed weeks
- no known major fetal malformations
- in active labour, but not in active second stage
- no known contra-indication to vaginal delivery
- clinical decision made to perform continuous CTG monitoring

Patients will not be included, if they are under 16 years of age or are not able to provide written informed consent.

DISSEMINATION OF THE STUDY TO POTENTIAL PARTICIPANTS

Patient study information (posters and leaflets) will be distributed to eligible women on their initial contact with healthcare professionals and revisited during antenatal and parentcraft classes, which take place at different stages of pregnancy. The same information will also be available to women on arrival in labour at participating hospitals.

CONSENT REQUEST AND ENROLMENT

Eligible women will be asked by their attending midwife, whether they wish to participate, when the clinical decision is made to perform continuous fetal monitoring during labour. The indication for instituting monitoring will be recorded. If receptive,
they will be asked for written informed consent and subsequently enrolled in the trial, in a window that is automatically opened by the Omniview-SisPorto® 3.5 program. If enrolment does not take place, the occurrence of exclusion criteria or the reasons reported by women for not participating will be registered.

RANDOMIZATION

After enrolment, women will be randomised to one of two arms using a 1:1 computer-generated randomisation sequence attributed to each centre by the Omniview-SisPorto® 3.5 program. The intervention arm will comprise continuous computer analysis of fetal monitoring signals with real-time alerts, while women allocated to the control arm will have continuous fetal heart rate monitoring as normally performed at that centre.
**INTERVENTION ARM**

Women randomised to the intervention arm will have continuous fetal monitoring during labour with computer analysis by the Omniview-SisPorto® 3.5 program with real-time alerts to changes detected on CTG and/or ST signals\(^\text{14}\). In this arm, ultimate management decisions remain the responsibility of healthcare professionals, according to their best clinical judgment. This may include CTG+ST analysis and/or FBS. However, non-directive guidelines are provided to help understand the meaning of the various alerts:

<table>
<thead>
<tr>
<th><strong>Signal loss/possible maternal heart rate monitoring</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- FHR with high signal loss or characteristics that are suggestive of maternal heart rate monitoring</td>
</tr>
<tr>
<td>- consider re-positioning the Doppler probe, changing to internal FHR monitoring, re-evaluating the fetal scalp electrode connections, or changing the electrode;</td>
</tr>
<tr>
<td>- ST signals with high signal loss</td>
</tr>
<tr>
<td>- consider re-evaluating the fetal scalp electrode connections or changing the electrode;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Tachysystole – excessive number of uterine contractions</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- consider discontinuing/reducing oxytocin infusion or consider acute tocolysis;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Yellow alerts</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- tracing characteristics that do not fulfill the criteria of normality, but are not usually associated with significant fetal hypoxia</td>
</tr>
<tr>
<td>- consider maintaining close monitoring and/or starting ST analysis if available;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Orange alerts</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- FHR+ST characteristics that may be associated with some degree of fetal hypoxia</td>
</tr>
<tr>
<td>- consider reversal of hypoxic causes if possible maintaining close monitoring, starting ST analysis if available, or performing FBS;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Red alerts</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- FHR+ST characteristics that are likely to be associated with fetal hypoxia</td>
</tr>
<tr>
<td>- consider immediate reversal of causes of hypoxia if possible or consider immediate delivery).</td>
</tr>
</tbody>
</table>
CONTROL ARM

Women randomised to the control arm will have continuous fetal monitoring during labour without computer analysis or alerts and will be managed according to the centre's existing guidelines. This may include ST analysis and/or FBS as adjuncts to standard CTG.

UMBILICAL CORD BLOOD ANALYSIS

All enrolled cases will be subject to immediate double cord clamping and paired cord blood sampling (umbilical artery and vein), which are prerequisites for accurate diagnosis of metabolic acidosis. Sampling into two pre-heparinised syringes should be delayed by no more than 30 minutes, air bubbles will be removed, the syringes capped, and blood analysis will occur in less than 30 minutes18.

COLLECTION OF BASELINE, LABOUR AND OUTCOME DATA

Baseline, labour and outcome data will be obtained by the local research midwife on the next working day after delivery and entered into the Omniview-SisPorto® 3.5 program. Baseline demographic and obstetric data of enrolled women will include maternal age, height, weight, pre-existing or current medical conditions, third trimester group B streptococcus carrier state (if known), number of previous pregnancies, number of previous vaginal and instrumental or caesarean deliveries. Data on the current labour will include gestational age, spontaneous or induced, normal or augmented, presence of slight or thick meconium staining of amniotic fluid, occurrence of fever (temp ≥38°C), use of epidural, parenteral or inhaled analgesia, other intrapartum medications, result(s) of FBS, date and time of birth, interval between end of fetal monitoring and delivery; normal, instrumental or caesarean delivery and indication for instrumental or caesarean delivery. Outcome data will include newborn birthweight, sex, 1 and 5-min Apgar scores, UApH (3 decimal places), pCO2, bicarbonate, BE_{ecf}, umbilical vein pH (3 decimal places), pCO2, bicarbonate, BE_{ecf}, NICU admission and indication. These data will be transmitted in an anonymised format to the co-ordinating centre. Cord acid base data will be assessed for accuracy. If cord blood samples are not available, neonatal data will be used to assess the occurrence of metabolic acidosis.
LATE OUTCOME AND SERIOUS ADVERSE EVENT DATA COLLECTION

Cases with metabolic acidosis (UApH < 7.05 and BE_{ecf} > 12 mmol/l), 5-minute Apgar score < 7 or NICU admission will be further investigated by the local research midwife to evaluate, whether neonatal blood analysis was performed in the first hour of life and its results including lactate, neonatal encephalopathy of any grade occurred in the first 72 hours of life, death of the infant occurred in the first 28 days of life, other important neonatal complications occurred in the first 7 days of life and brain ultrasound or other imaging technologies were performed in the first 7 days of life and their results. At the same time, serious adverse events (see data safety monitoring committee, below) will be sought and reported.

PRIMARY OUTCOME MEASURE

The primary outcome measure of this study is the incidence of fetal metabolic acidosis (defined as an UApH < 7.05, BD_{ecf} > 12 mmol/l). BD_{ecf} will be calculated from pH and PCO_{2} values, according to the Siggaard-Andersen acid-base algorithm. 17

SECONDARY OUTCOME MEASURES

The secondary outcomes are: overall rates of caesarean section and of caesarean section for non-reassuring fetal state, overall rates of instrumental vaginal delivery and of instrumental vaginal delivery for non-reassuring fetal state, FBS rates, incidence of 5-minute Apgar score < 7, need for NICU admission, incidence of moderate and severe neonatal encephalopathy with a hypoxic marker; perinatal death, rate of delayed interventions (interval between red alerts [offline analysis in the control arm] and delivery in metabolic acidosis cases), internal FHR monitoring rates, tracing quality and signal loss.

SAMPLE SIZE

The sample size calculation is based on the primary endpoint: a metabolic acidosis incidence of about 2.8% can reasonably be assumed, as this was previously reported in an observational study conducted in one of the participating centres11. Since this will be the first trial to evaluate the effect of computer analysis of intrapartum fetal monitoring signals and real-time alerts on perinatal outcomes, it is not possible to find an estimate of the degree of change that is expected. The closest available parallel is the evaluation of STAN® monitoring versus conventional CTG during labour. A systematic review of
the first three trials, that studied this issue, revealed an overall relative risk for metabolic acidosis of 0.64. In the absence of a better alternative, this will be the value used for initial sample size calculation. Thus assuming a reduction in metabolic acidosis from 2.8% to 1.8%, with an $\alpha$ of 0.05, a two-sided test, and a power of 0.80, about 7320 women will need to be randomised. Accounting for a 10% loss to follow-up, the study requires the inclusion of 8133 women in order to obtain the 7320 analysable cases (3660 per arm). A pilot analysis will be conducted after enrolment of the first 1500 cases to evaluate the real incidence of metabolic acidosis, the effect of real-time alerts on its rate and the need to recalculate the trial’s sample size.

DATA ANALYSIS

Data analysis will be carried out at the co-ordinating centre and analysis of the primary endpoint will follow the intention to treat principle. Minimal differences between groups are expected in baseline patient characteristics. The incidence of metabolic acidosis will be compared across both groups, using RR with $95\%$CI. A similar methodology will be applied to secondary outcomes.

MISSING DATA

A case will be classified as having metabolic acidosis, if it falls within the described diagnostic criteria and the arterial sample shows a pH value at least 0.03 units lower and a PCO$_2$ value at least 1kPa higher than the venous sample$^{17,18}$. If these two latter criteria are not met or only one sample is available, it will be considered as indicating a venous sample. Venous samples will still be considered for the diagnosis of metabolic acidosis, if it meets the diagnostic criteria. If no cord blood acid base data is available, the case may still be classified as having metabolic acidosis, if the neonate has acid-base data obtained in the first hour of life that fulfill the diagnostic criteria, or if a lactate value greater than 10 mmol/l is documented in association with signs of maladaptation (respiratory distress syndrome, need for buffering, etc).

LOCAL RESEARCH CO-ORDINATORS AND LOCAL RESEARCH MIDWIVES

A research co-ordinator in each centre will be responsible for planning of the trial and for dissemination of information to staff. A local research midwife will contribute to these tasks and be responsible for dissemination of information to possible participants, encouraging high levels of patient recruitment, regular and timely
collection and recording of data and provision of regular feedback to the co-ordinating centre on study progress and potential protocol violations.

ETHICAL CONSIDERATIONS

The study protocol will be evaluated by the relevant UK research ethics committees. Written informed consent from all participants will be obtained before trial enrolment.

This is the first time that computer analysis of intrapartum fetal monitoring with real-time alerts is being compared with conventional fetal monitoring in a randomised trial, so there is no prior evidence of benefit for the reduction in the incidence of adverse neonatal outcomes. However, there are observational data to suggest that computer analysis of fetal monitoring signals has a higher validity in prediction of adverse outcomes16.

The confidentiality of personal data is guaranteed by the fact that person-identifiable data will not be available to anyone outside the local healthcare team. All data transmitted to the co-ordinating centre will be anonymised automatically.

INDEPENDENT DATA SAFETY MONITORING COMMITTEE

All serious adverse events (see below) will be reported to the Data Safety Monitoring Committee, who will evaluate their frequency regularly and determine whether there is a significantly increased incidence in the intervention group and if so, whether the study should be discontinued.

This committee will consist of a neonatologist, an obstetrician and a statistician. Serious adverse events are defined as any of the following:

- severe metabolic acidosis (UApH < 7.00 and BE_{eef} > 12 mmol/l) and NICU admission;
- 5-minute Apgar score < 7 and NICU admission;
- First available pH value after birth < 7.05 or first available lactate value after birth > 10 mmol/l;
- Grade II or III neonatal encephalopathy;
- Death in the first 28 days of life.
REFERENCES

Chapter 8

General Discussion

Main Conclusions
GENERAL DISCUSSION

Fetal monitoring during labour constitutes a challenge for human information management. Until recently, the labour and delivery nurse/midwife/obstetrician managed this complex situation by visual analysis of a host of information. Modern developments have identified the possibility of not only adding new information from ST analysis, but also applying computer assisted CTG+ST data interpretation with the potential of reducing ambiguity and improving coherence within clinical guidelines.

To achieve a paradigm shift from a screening to a diagnostic capacity continuous fetal monitoring during labour requires not only additional information, but also a more objective and systematic management of existing data.

The SisPorto® system has seen a long line of development, utilising a combination of engineering and clinical skills initially focusing on CTG assessment, but lately adding the output of ST waveform analysis generated by a dedicated data acquisition/ signal processing unit – the STAN® fetal heart recorder.

The main objective of this thesis was to develop and evaluate a methodology, which integrates computerised analysis of CTG and ST event features, with the aim of improving the sensitivity and specificity of intrapartum fetal monitoring in prediction of neonatal outcome. In order to accomplish this objective, the following research phases were established:

1. Optimisation of the Omniview-SisPorto® CTG analysis algorithms and integration with fECG ST event data with the aim of developing computerised standards for real-time clinical alerts.
2. Comparison of the system’s analysis of basic CTG features: baseline, accelerations, decelerations and contractions to that of a consensus of experts.
3. Evaluation of the impact of clinicians’ access to computerised analysis of CTGs on their reproducibility and accuracy in prediction of neonatal outcome.
4. Evaluation of the Omniview-SisPorto® alerts in prediction of newborn UA acidemia.
5. Planning of a RCT to evaluate the effectiveness of the use of Omniview-SisPorto® in the reduction of adverse maternal and neonatal outcomes.
OPTIMISATION OF THE OMNIVIEW-SISPORTO®

The Omniview-SisPorto® 3.5 program (Speculum, Lisbon, Portugal) is the most recent version of a central monitoring system, that provides visual and sound alerts, based on computer analysis of both CTG and ST event features. Chapter 3 describes the program's main characteristics and provides an overview of the system’s online alerts. Omniview-SisPorto® 3.5 is the first central monitoring system to incorporate computerised analysis of CTG and ST event features, providing healthcare professionals with real-time alerts for minor and major changes in monitored signals. This system has been installed in more than 20 hospitals worldwide and has been tested in over 20 000 pregnancies.

COMPARISON OF THE SYSTEM’S ANALYSIS OF BASELINE, ACCELERATIONS, DECELERATIONS AND CONTRACTIONS WITH THAT OF A CONSENSUS OF EXPERTS.

Chapter 4 describes a study that evaluated the agreement between visual and computer analysis of intrapartum CTG features. A high agreement was found between Omniview-SisPorto® 3.5 and a consensus of experts in the assessment of baseline heart rate and on identification of accelerations, decelerations and contractions.

Strengths and limitations

The large number of FHR sequences evaluated in this study and the use of strict timing criteria to identify specific CTG features gives strength to these results. The interpretation of agreement on identification of accelerations and decelerations is much more uncertain than that of the baseline, due to the higher interobserver disagreement found in evaluation of these parameters. Nevertheless, it is reassuring to observe that the computer agrees with the consensus analysis as much as do individual clinicians. The results obtained in the present study suggest a higher agreement on identification of accelerations and decelerations to that obtained with other systems. It became clear during data analysis that periodic events with an amplitude and/or duration close to the cut-off values were the most frequent cause of disagreement. This occurred both in interobserver analysis and in the computer-consensus comparison. Agreement on the identification of periodic events is particularly sensitive to small disagreements on
baseline estimation, which can by itself lead to divergence over the former. The clinical significance of disagreement over the identification of accelerations and decelerations of such small amplitude is probably very limited, at least in term fetuses.

The answer to the question: “How does computer evaluation of basic CTG features compare with that of experts” was accomplished with satisfactory results, although further improvement in some of the algorithms can still be envisioned. This issue is of the utmost relevance, as it is unlikely that a computer system will be welcome in clinical practice, if healthcare professionals working with it do not relate to the analysis of basic CTG features.

On the other hand, agreement studies between visual and computerised CTG analysis are forcibly of limited value, because computer systems were developed with the primary aim of overcoming the well-demonstrated poor intra- and interobserver agreement on visual analysis. However, this comparison is unavoidable, due to the absence of a gold standard for FHR interpretation.

Further research related to this topic should probably include the evaluation of agreement regarding the classification of cardiotocographic tracings, with and without ST data, and agreement on clinical management. The latter could include the evaluation performed by clinicians with and without access to cCTG analysis with or without ST waveform analysis. Automated ST analysis requires a steady-state situation at onset (a stable baseline and a non-deteriorating FHR pattern), as well as continuous signs of a fetus capable of responding (signs of reactivity and variability). Provided these pre-requisites are fulfilled, access to ST analysis may reduce the ambiguity in CTG assessment, as classification is mainly required in the case of ST events being present, focusing on the strength of the CTG as an indicator of normality.
This study, described in Chapter 5, demonstrates that clinicians achieve a higher agreement on prediction of UApH and are more accurate in this prediction, when they have access to computerised analysis of CTG tracings. This suggests that they are influenced by the results of computer analysis, and that perhaps this leads to a more homogeneous and precise interpretation of tracings.

Strengths and limitations

One may consider that a correct prediction of newborn UApH values within a 0.10 margin in 70% of tracings is not an impressive result, particularly given that the standard deviation of this value was 0.08. Nevertheless, the accuracy of prediction was significantly higher in the group that had access to computer analysis. Furthermore, access to cCTG results was associated with the prediction of less acidaemia. The most likely explanation for this is that the information conveyed by the cCTG analysis adds to observers’ ability to identify criteria of normality. A much larger sample size would be needed in order to narrow the tolerance window around the UApH values or, alternatively, an artificial selection of cases with adverse neonatal outcome would need to be performed. A higher number of cases would also allow a sensitivity and specificity approach to the results of this study.

Evaluating the degree of fetal acidemia, rather than identifying dichotomous classes of UApH (acidemic, non-acidemic) seems nevertheless to be a clinically useful option, as this is estimation is frequently employed, when deciding the timing of a clinical intervention, which ideally should be before the onset of metabolic acidosis.

Care was taken to obviate methodological bias by excluding cases with large signal loss and keeping the time-lag between the end of FHR monitoring and delivery to a minimum, so that tracings reflected as much as possible fetal oxygenation at the time of delivery.

Prediction of Apgar scores from CTG tracings has a much more limited value, because of the inherent subjectivity and poor reproducibility in its evaluation and because of the influence of factors other than fetal oxygenation.

The results of this study suggest that access to cCTG analysis has the potential to improve clinical decision-making as a result of a more consistent and precise evaluation of CTG features. Whether this will reduce the incidence of adverse neonatal outcomes, needs to be adequately evaluated in a RCT.
EVALUATION OF THE OMNIVIEW-SISPORTO® SYSTEM’S ACCURACY IN PREDICTION OF NEWBORN UA ACIDEMIA.

This study, described in Chapter 6, demonstrated that the Omniview-SisPorto® system achieved a high accuracy in prediction of UA neonatal acidemia (94.5%), and detected all cases of severe neonatal acidemia.

Strengths and limitations

Although the confidence intervals for sensitivity were still wide, due to limited number of index cases, all acidic cases were identified, highlighting the need for further studies to confirm this result. On the other hand, the sensitivity and the positive likelihood ratio were very high, suggesting that the system will help to improve clinical judgement on neonatal outcome.

Establishment of strict inclusion and exclusion criteria to avoid potential biases led to the selection of a population at high risk of adverse neonatal outcome. This selection has advantages, but also some limitations. The positive aspect is that the population is more prone to adverse neonatal outcome, which is the primary outcome of the study. A limitation is that, by excluding other pregnancies that are at lower risk of neonatal acidemia, the system is not being tested in a general population. The objective of a central monitoring station with automated CTG and ST analysis is to serve the general obstetric population. The method’s positive and negative predictive values depend on the prevalence of the outcome, but this is not the case with sensitivity and specificity.

An important issue to remember is that this was an observational study, where computer analysis was compared a posteriori with neonatal outcome. This evaluation does not take into account clinician management decisions, which can influence neonatal outcome.

The chosen cut-off value for defining neonatal acidemia as a pH<7.05 is the most widely used in current research, but is far from consensual. It would also be interesting to include metabolic acidosis and hypoxic-ischemic encephalopathy as outcome measures, but the study’s sample size was insufficient for this.

In this study only the alerts elicited by the system during the last 60 minutes of the tracing were correlated with UApH. Neonatal outcome is well-known to be influenced by processes that last more than 60 minutes or may even begin before labour. However, to test the accuracy of all alerts displayed during the course of labour, the most related outcome would probably be the results of FBS, as rapid changes in fetal acid-base state can occur during labour. FBS is known to have limitations, so its use as a “gold standard” for this study would be highly questionable.
EVALUATION OF THE EFFECTIVENESS OF THE OMNIVIEW-SISPORTO® SYSTEM IN REDUCTION OF ADVERSE OUTCOMES.

The ideal study to assess the effectiveness of the Omniview-SisPorto® system in reduction of adverse neonatal and maternal outcomes is a randomised controlled trial. This evaluation needs to take place in real labour ward settings, taking into account the effect of clinical management decisions based on the system’s alerts. Chapter 7 describes the study design of this RCT in detail. The following hospitals have agreed to participate: St. George’s Hospital University of London; Cardiff University Hospital, Wales; Ninewells Hospital, Dundee, Scotland; Glan Clwyd Hospital, Rhyl, Wales. Patient recruitment is scheduled to begin in October 2010.
MAIN CONCLUSIONS

✓ Omniview-SisPorto® is the first central monitoring system to incorporate computerised analysis of the CTG and ST event features, providing real-time visual and sound alerts for features associated with poor fetal oxygenation.

✓ A high agreement was obtained, when comparing the evaluation of basic cardiotocographic features by Omniview-SisPorto® and a consensus of clinicians.

✓ Access to computerised analysis of intrapartum CTGs performed by Omniview-SisPorto® improved clinician’s agreement and accuracy in prediction of newborn umbilical pH.

✓ Computer analysis of FHR and ST event signals performed by Omniview-SisPorto® obtained a sensitivity of 1.00 95CI[0.56;1.00] and a specificity of 0.94 95CI[0.89;0.97] in prediction of UA acidemia. The system identified all cases of severely acidemic fetuses.

✓ This method of intrapartum fetal monitoring will continue to be refined over the coming years. The effectiveness of the system will be evaluated in a multicentre RCT, starting in October 2010.

In 1996 Herbert van Geijn wrote:”...there is still a long way to go until decision support systems find their way into obstetric practice”. Since then, new computer systems have been invented, some have been improved, and some have been abandoned, new adjuncts to CTG monitoring have been developed, and extensive research on this subject has been conducted. Integration of these systems into routine clinical practice is finally a reality. The forthcoming clinical trials, aimed at evaluating their effectiveness, should throw further light into the future of a much awaited computer-assisted fetal monitoring during labour and delivery. Perhaps the time has finally come for this paradigm shift…
List of publications
This PhD thesis is based on the following scientific publications:


- Costa A, Machado AP, Costa-Santos C, Ayres-de-Campos D, Bernardes J. **Comparison of the evaluation of cardiotocographic events by Omniview-SisPorto 3.5 and a consensus of clinicians.** *J Perinat Med, 2010,* [Epub ahead of print.]


Acknowledgments

Agradecimientos
Numerosas pessoas contribuíram, directa ou indirectamente, e por vezes não o sabendo, para esta tese. A todas elas a minha mais sincera gratidão.

**Quero agradecer a:**

**Prof. Diogo Ayres-de-Campos**, meu orientador, que me introduziu nesta área da monitorização fetal e que, acima de tudo, acreditou em mim.

**Prof. Karl Rosén**, my co-supervisor, for sharing his vast knowledge and experience in fetal physiology and ECG analysis, as well as his invaluable suggestions to the present thesis.

**Prof. Bernardes**, que me serve de referência nesta área tão importante da monitorização fetal, mas também pela sua valiosa experiência na área da reprodutibilidade e validação científica na medicina, pelas suas sempre pertinentes críticas, que me fazem crescer.

**Mestre Cristina Santos**, não só pela sua preciosa mais-valia na análise estatística dos dados, mas também pela disponibilidade, que sempre manifestou.

**Prof. Nuno Montenegro**, pelo seu apoio incondicional.

**Dr. Vera Paiva**, pela sua amizade sem limites, que esteve sempre comigo nos bons momentos e mesmo naqueles mais difíceis. Obrigada, Vera!

**A todos os co-autores dos artigos da tese**, sem os quais esta tese não seria possível.

**D. Célia Teixeira**, pelo fornecimento de artigos científicos (eu sei que a tarefa não foi fácil!)

**A todos os meus colegas de trabalho do Departamento de Ginecologia e Obstetrícia** pelo apoio e amizade.

Finalmente, a toda a minha família e amigos, obrigada por existirem e me fazerem existir…
Dissertação de candidatura ao grau de Doutor em Medicina, na área de Obstetrícia, submetida à Faculdade de Medicina da Universidade do Porto.