Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Obstetrics & Gynecology and Reproductive Biology 118 (2005) 52–60
EUROPEAN JOURNAL OF OBSTETRICS & GYNECOLOGY AND REPRODUCTIVE BIOLOGY
www.elsevier.com/locate/ejogrb

Prediction of neonatal state by computer analysis of fetal heart rate tracings: the antepartum arm of the SisPorto® multicentre validation study

Diogo Ayres-de-Campos[a,*], Cristina Costa-Santos[b], João Bernardes[a]
For the SisPorto® Multicentre Validation Study Group[1]

[a]*Departamento de Ginecologia e Obstetrícia, Faculdade Medicina da Universidade do Porto, Alameda Hernani Monteiro, 4200-319 Porto, Portugal*
[b]*Departmento de Bioestatístíca e Informática Médica, Faculdade de Medicina da Universidade do Porto, Alameda Hernani Monteiro, 4200-319 Porto, Portugal*

Received 2 February 2004; received in revised form 4 March 2004; accepted 13 April 2004

Abstract

Objective: To evaluate the capacity of computer analysis of antepartum cardiotocographs performed by SisPorto® 2.0 in predicting neonatal outcome. *Study design*: A prospective observational study was conducted in eight tertiary care centres in Europe and Australia, involving pregnant women in the absence of labor, scheduled for elective caesarean section, whose last fetal heart rate (FHR) tracing was performed within 4 h of delivery. After exclusion of fetal malformations, multiple pregnancies, tracings with less than 30 min, tracings with more than 15% signal loss, difficult fetal extractions, and anesthesia complications, a total of 345 cases were analyzed. Computer quantification of cardiotocographic parameters was compared with newborn Apgar score, umbilical artery pH, metabolic acidosis and neonatal hypoxic-ischemic encephalopathy, by means of receiver operating characteristic (ROC) curves. *Results*: Acceleration number, mean short-term variability, percentage of abnormal short-term variability and percentage of abnormal long-term variability had an excellent discriminative capacity to predict 1-min Apgar scores under or equal to 4 (areas under the ROC curve 0.96–1.00). The same parameters showed a slightly lower capacity to predict 5-min Apgar scores under or equal to 6 (areas under the ROC curve 0.81–0.89). The best cut-off values for these parameters, derived from the previously referred calculations, detected all cases of hypoxic-ischemic encephalopathy ($n = 2$). Cardiotocographic parameters showed a lower discriminative capacity in prediction of umbilical artery pH <7.20 (maximum area under the ROC curve 0.66) and <7.15 (maximum area under the ROC curve 0.69). *Conclusions*: Computerized quantification of accelerations and variability in the antepartum allows a good prediction of 1 and 5-min Apgar scores, and to a much lesser degree umbilical artery pH.

Keywords: Fetal heart rate; Cardiotocography; Neonatal outcome; Validity; Computer analysis

* Corresponding author. Tel.: +351-966707112; fax: +351-225096068.
E-mail address: sisporto@med.up.pt (D. Ayres-de-Campos).

[1] The SisPorto® Multicentre Study Group involved the following investigators and centers: Dr. Bernd Klein from the Aachen University Hospital, Germany; Drs. John Svigos, Julia Dalton and Jan Hill from the Calvary Hospital, Adelaide, Australia; Drs. Philip Banfield and Adelaja Mustapha from the Glan Clwyd Hospital, Rhyl, Wales; Prof. Sophie Alexander and Dr. Christine Kirkpatrick from the Hospital Erasme, Université Libre de Bruxelles, Belgium; Dr. Isabel Fagulha and Prof. Paulo Moura from the Hospitais da Universidade de Coimbra, Portugal; Prof. Manuel Meirinho, Drs. Rui Ribeiro and Maria Delgado from the Hospital Garcia de Orta, Almada, Portugal; Prof. Luís Graça, Drs. Nuno Clode and Miguel Tuna from the Hospital de Santa Maria, Lisbon, Portugal; Prof. Santos Jorge, Drs. Ricardo Marques†, Teresa Coelho, Graça Buchner, and Manuela Sequeira from the Hospital Santo António, Porto, Portugal; Prof. Luis Cabero and Dr. Maria Cerqueira from the Hospital Vall d'Hebron, Barcelona, Spain; Drs. Carsten Nickelsen and Lars Meinard-Jensen from the Hvidovre University Hospital, Copenhagen, Denmark; Prof. Gunvor Ekman, Drs. Lena Granstrom and Jan Rapp from the Karolinska Institute, Stockholm, Sweden; Dr. Leo Mäkäräinen from the Oulu University Hospital, Finland; Prof. Wolfgang Holzgreve and Dr. Carolyn Troeger from the University Women's Hospital, Basel, Switzerland; and Drs. Ana Matos, Joaquim Saraiva and Ana Reynolds from the São João Hospital, Porto, Portugal. Study design and coordination: Profs. Diogo Ayres-de-Campos, João Bernardes and Luís Pereira-Leite. Statistical analysis: Dr. Cristina Costa-Santos and Prof. Altamiro Costa-Pereira. Technical support: Eng. António Garrido and Prof. Marques-de-Sá.

0301-2115/$ – see front matter
doi:10.1016/j.ejogrb.2004.04.013

1. Introduction

More than 30 years after the introduction of antepartum cardiotocography into clinical practice, the predictive capacity of the method remains controversial. In a review of 45 articles published on this subject, Devoe et al. found that its reported sensitivity varies between 2 and 100%, and its specificity between 37 and 100% [1]. Similar results are reported in other reviews [2,3]. There are many possible explanations for these discrepant numbers. Different cardiotocographic equipment was used in the studies, as well as different criteria for interpreting fetal heart rate (FHR) tracings, and varying definitions of poor neonatal outcome. Many studies were performed before important aspects of FHR analysis started to be taken into account, such as FHR variability and the existence of fetal behavioral states. Another important issue is the time interval that elapsed between the end of the study tracing and evaluation of neonatal state at delivery. In some studies this interval exceeded 1 week, in some it is not revealed, and in others it comprises situations capable of inflicting a significant stress to the fetus, such as labor. Intervention bias can also be problem, in what is known as the ''treatment paradox'': when FHR tracings associated with fetal stress, but not distress, are considered a ''positive test'' and an intervention is performed, a high number of false positives may occur, because poor outcome was effectively avoided by intervention. Finally, interpretation of these results must take into account the well-demonstrated poor reproducibility of visual analysis [4].

Randomized trials on antepartum cardiotocography have failed to show a difference in the incidence of low Apgar scores, elective caesareans, induced labours, neonatal neurological outcome, intensive care unit admissions or perinatal mortality [5]. However, it must be emphasised that all these trials were conducted in the early 1980's when other methodologies for fetal evaluation were common practice, they evaluated a relatively small number of cases, and used very different FHR monitoring intervals. A detailed analysis of clinical records in the few cases of fetal death showed that mortality in both groups could largely be explained by intrapartum events [6].

Computer analysis of cardiotocographs has the theoretical advantage of providing a reproducible and objective interpretation of FHR tracings, quantifying parameters that are difficult to assess by the human eye, such as short- and long-term variability. Such systems have been developed since the late 1970's [7] but have still to gain wide clinical acceptance. This may be related to the limited demonstration of their validity and efficacy, as well as to the poor practicality of their routine clinical use. SisPorto® 2.0 is a program for automated analysis of tracings, developed over the last 14 years at the University of Porto. FHR baseline is calculated using a complex algorithm developed to identify the mean FHR during stable segments, in the absence of fetal movements and uterine contractions [8]. Accelerations are defined as increases in the FHR above the baseline, lasting 15–120 s and reaching a peak of at least 15 bpm. Decelerations are defined as decreases in the FHR under the baseline, lasting at least 15 s and with an amplitude exceeding 15 bpm. They are classified as mild if shorter than 120 s, prolonged if they last 120–300 s, and severe if they exceed 300 s. Points with abnormal short-term-variability (STV) are recognised when the difference to adjacent FHR signals is less than 1 bpm. Points with abnormal long-term variability (LTV) are identified whenever the difference between maximum and minimum FHR values of a sliding 60-s window centered on them, does not exceed 5 bpm. The system is extensively described elsewhere [8].

The main objective of this study was to evaluate the capacity of SisPorto® 2.0's FHR parameters to predict various indicators of neonatal outcome (1 and 5-min Apgar score, umbilical artery pH, metabolic acidosis and neonatal hypoxic-ischemic encephalopathy). The least possible interval between FHR monitoring and newborn evaluation was aimed at, with further exclusions of cases where situations associated with fetal stress occurred during this interval.

2. Materials and methods

A prospective multicentre observational study was conducted in which all tracings were acquired with Doppler ultrasound using Sonicaid® Meridien 800, Hewlett-Packard® M1350A, M1350B or M1351-3 fetal monitors, connected by cable to a personal computer running the SisPorto® 2.0 program. Necessary criteria for inclusion were the acquisition of tracings with at least 30 min duration and less than 15% signal loss. Exclusion criteria were multiple gestations, and fetal malformations diagnosed before or after birth. In order to minimize the biases that were referred to previously, only women scheduled for elective cesarean section in the absence of labor, whose FHR tracings were acquired within the 4 h that preceded delivery, were included. The study population thus consisted mainly of in-patients from high-risk obstetric wards of tertiary care hospitals, with an established indication for elective cesarean section, based on criteria that did not take into account the results of computer FHR analysis. Women operated under general and regional anesthesia were included, as there is no evidence that this significantly influences neonatal outcome [9,10]. On the other hand, difficult fetal extractions during cesarean section ($n = 2$) and anesthesia complications occurring before or during surgery ($n = 1$) were excluded, as these have been associated with neonatal acidemia [10,11]. This information was obtained from the Obstetricians and Anesthesiologists involved in the procedures.

In each case, the last available FHR tracing lasting 30–60 min was evaluated, and the following parameters were extracted from the SisPorto® 2.0 analysis: FHR baseline, number of accelerations, percentage of the tracing with

abnormal short-term variability, average short-term variability, percentage of the tracing with abnormal long-term variability, average long-term variability, and number of mild, prolonged, and severe decelerations.

Apgar scores at 1, 5, and 10 min were evaluated by the health caregivers responsible for immediate neonatal support in each participating center. The cut-off values considered clinically most meaningful were selected: 1-min Apgar score ≤4, 5-min Apgar score ≤6, and 10-min Apgar score ≤6. For evaluation of umbilical blood gas values, the cord was doubly clamped immediately after birth and within 30 min blood was aspirated from both artery and vein into two previously heparinized syringes. After vestigial air was expelled and the needle protected, blood gas analysis was carried out within 30 min using the equipment available in each center [12,13]. Cases with no data on either arterial and venous samples, and those in which pH values differed by less than 0.03 units, or PCO_2 values differed by less than 7.5 mm Hg, were excluded from the analysis of this parameter [14]. Cut-off values for arterial pH at 7.15 and 7.20 were evaluated. Metabolic acidosis was defined as an umbilical artery pH <7.05 and an extracellular fluid base deficit exceeding 12 mmol/l. In newborns with arterial pH <7.10, 1-min Apgar score ≤4, or 5-min Apgar score ≤6, ($n = 11$) hospital records were reviewed for the diagnosis of hypoxic-ischemic encephalopathy. This was defined as the appearance of changes in muscle tone, feeding, state of conscience or seizures occurring in the first 48 h of life.

A total of 345 cases were included in the study. The number of cases included and excluded in each participating center is displayed in Table 1. Ethical committee approval for the study and informed consent for enrollment of subjects was obtained in all participating centers.

2.1. Statistical methods

The capacity of each FHR parameter to predict neonatal outcome indicators was evaluated by means of receiver operating characteristic (ROC) curves [15]. Areas under the ROC curve were calculated with 95% confidence intervals (CI). The best cut-off value was chosen as the one with the highest sensitivity and specificity. For calculation of statistical data SPSS for Windows® version 10.0.7 was used.

3. Results

Maternal age varied between 16 and 43 years, with an average of 30 and a standard deviation of 5. Risk factors for fetal hypoxia in the study population are displayed in Table 2. The majority of cases (60.9%) had no identifiable risk factors. Median gestational age was 38 weeks, with a minimum of 25 and maximum of 42. One woman had a gestational age of less than 28 weeks, in five it was between 28 and 30 weeks, in 10 between 30 and 32 weeks and in 16 cases between 32 and 34 weeks.

Newborns weighed between 750 and 4570 g, with an average of 3023 g and a standard deviation of 801. Fifty-one percent were female and 49% male. Endotracheal intubation in the delivery ward was carried out in 22 cases (6%) and there were 53 neonatal intensive care unit admissions (15%). Two neonatal deaths occurred: one on the 4th day after birth caused by a hyperglycemic coma and the other on the 7th day due to sepsis of a probable intrauterine origin.

Eight newborns were attributed a 1-min Apgar score ≤4, six a 5-min Apgar score ≤6 and two a 10 min Apgar score ≤6. Of the 270 cases with valid umbilical blood gas values, 46 had an arterial pH <7.20, 15 a pH <7.15, two a pH <7.10 and one a pH <7.05. There were no cases of metabolic acidosis. In two newborns the diagnosis of hypoxic-ischemic encephalopathy was established; one of them grade II and the other grade III with subsequent neonatal death on the 7th day of life. The Apgar scores of these two cases were 3/6/7 and 4/8/8, respectively. In the first case umbilical blood samples were not collected, and in the second an arterial pH of 7.15 and a base deficit of 11 mmol/l were recorded.

The small number of cases with 10-min Apgar score ≤6 ($n = 2$), hypoxic-isquemic encephalopathy ($n = 2$) and metabolic acidosis ($n = 0$) precluded a formal evaluation of the validity of FHR parameters in the prediction of these outcomes. The number of valid neonatal parameters, their median, 2.5 and 97.5 percentiles, minimum and maximum values are considered in Table 3.

Table 1
Number and source of included and excluded cases

Participating centers	Included	Excluded
Aachen University Hospital, Germany	39	2
Calvary Hospital, Adelaide, Australia	18	3
Glan Clwyd Hospital, Rhyl, United Kingdom	23	6
Hospitais da Universidade de Coimbra, Portugal	26	8
Hospital Santo António, Porto, Portugal	50	8
Hospital São João, Porto, Portugal	151	35
Hospital Vall d'Hebron, Barcelona, Spain	–	24
University Women's Hospital, Basel, Switzerland	38	2
Total	345	88

Table 2
Risk factors for fetal hypoxia in the study population

	Number
Maternal diabetes	22
Other severe maternal health conditions	14
Pre-eclampsia without detected fetal compromise	28
Fetal growth restriction, oligohydramnius and/or altered diastolic umbilical blood flow	53
Rh immunization	3
Isolated cardiotocographic abnormalities	5
Placenta previa	10

Table 3
Main characteristics of neonatal and FHR parameters in the study population

	n	Median	Percentile 2.5	Percentile 97.5	Minimum	Maximum
1-min Apgar score	342	9	5	10	1	10
5-min Apgar score	342	10	7	10	4	10
10-min Apgar score	190	10	9	10	5	10
Arterial pH	270	7.27	7.11	7.35	7.05	7.37
FHR baseline	345	130	110	149	110	171
Acceleration number	345	13	0	37	0	47
Percentage of abnormal STV	345	53	29	75	25	88
Mean STV	345	1.1	0.5	4.1	0.3	11.9
Percentage of abnormal LTV	345	2	0	37	0	95
Mean LTV	345	8.7	1.7	19.6	0.4	23.5
Prolonged decelerations	345	0	0	0	0	2

STV: short-term variability, LTV: long-term variability.

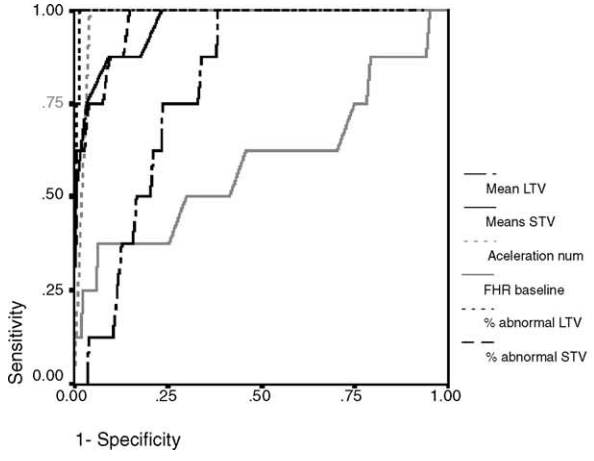
Accelerations occurred in 319 tracings (92%) and prolonged decelerations (2–5 min) in 3 (1%). No tracings demonstrated severe (>5 min) or repetitive decelerations [6]. Other characteristics of the distribution of FHR parameters are considered in Table 3. Analysis of the different types of decelerations is not included in the ROC figures, as it failed to provide satisfactory discriminative capacities.

3.1. Prediction of 1-min Apgar score ≤4

Both cases that went on to develop hypoxic-ischemic encephalopathy were included in this group. The capacity of FHR parameters to predict this outcome is displayed in Fig. 1, by means of ROC curves and areas under these curves. A scatter graph comparing the distribution of 1-min Apgar score with the percentage of abnormal LTV is shown in Fig. 2.

The FHR parameters with best discriminative capacities (accelerations, percentage of abnormal short-term variability and percentage of abnormal long-term variability) were also analyzed in combination, using the best cut-off values arrived at by individual analysis. Tracings without accelerations were evaluated to assess whether the quantification of abnormal short and long-term variability would discriminate between cases with low and high 1-min Apgar scores. In Fig. 2 a scatter graph relating abnormal long-term variability with 1-min Apgar scores in the 26 cases without accelerations is shown.

Absence of accelerations combined with abnormal long-term variability >30.5% obtained a sensitivity of 100% (95%CI: 63–100) and a specificity of 99% (95%CI: 97–100) in predicting 1-min Apgar score ≤4. Three false positives and no false negative occurred in this evaluation. The area under the ROC curve for this evaluation was 0.997 (95%CI: 0.992–1.000. When absence of accelerations was combined with abnormal short-term variability >63.5%, sensitivity was 100% (95%CI: 63–100) and specificity 98% (95%CI: 95–99). Abnormal long-term variability >30.5% combined with abnormal short-term variability >63.5% obtained a sensitivity of 100% (95%CI: 63–100) and a specificity of 99% (95%CI: 97–100). The combination of all three parameters provided a sensitivity of 100% (95%CI: 63–100) and a specificity of 97% (95%CI: 95–99). If FHR baseline above 150 bpm was combined with the



FHR parameters	Area	95% CI		Best
		Lower	Upper	
Acceleration num.	0.979	0.964	0.994	0.5
% abnormal STV	0.966	0.930	1.000	63.5
Mean STV	0.961	0.914	1.000	0.65
% abnormal LTV	0.996	0.990	1.000	30.5
Mean LTV	0.798	0.716	0.881	8.05
FHR baseline	0.594	0.349	0.838	143.5

Fig. 1. ROC curves for the prediction of 1-min Apgar ≤4 ($n = 8$) by the studied FHR parameters. On the right, areas under the ROC curves, 95% confidence intervals, and best cut-off values for each parameter. STV: short-term variability, LTV: long-term variability.

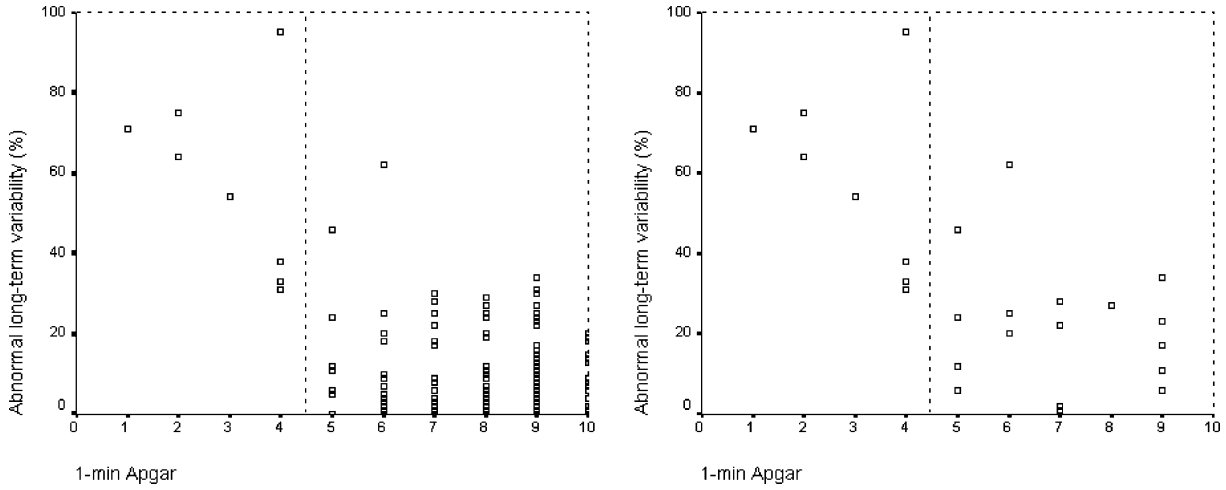


Fig. 2. Scatter-graphs relating the percentage of abnormal long-term variability with 1-min Apgar scores in the whole study population (on the left) and in cases without FHR accelerations (on the right, $n = 26$).

absence of accelerations and abnormal long-term variability >30.5%, sensitivity decreased to 12% (95%CI: 1–53), but specificity remained at 99% (95%CI: 98–100). None of the cases with low 1-min Apgar score exhibited prolonged, severe or repetitive decelerations.

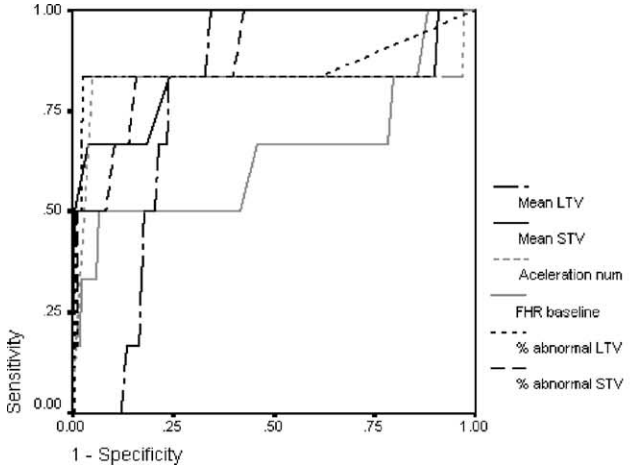
3.2. Prediction of 5-min Apgar score ≤6

Five of the six cases with 5-min Apgar score ≤6 also had a 1-min Apgar score ≤4. The ROC curves for the prediction of this outcome by FHR parameters are displayed in Fig. 3.

Combinations of FHR parameters were also evaluated, using the best cut-off values obtained with individual analysis. Absence of accelerations combined with abnormal long-term variability >30.5% provided a sensitivity of 83% (95%CI: 36–100) and a specificity of 98% (95%CI: 96–99) in prediction of this outcome. Absence of accelerations together with abnormal short-term variability >63.5%, obtained a sensitivity of 83% (95%CI: 36–100) and a specificity of 97% (95%CI: 94–98). If abnormal long-term variability >30.5% was combined with abnormal short-term variability >63.5%, the sensitivity was 83% (95%CI: 36–100) and the specificity 98% (95%CI: 96–99). The combination of all three parameters provided a sensitivity of 83% (95%CI: 33–100) and a specificity of 96% (95%CI: 94–98). If FHR baseline above 150 bpm was combined with the absence of accelerations and abnormal long-term variability >30.5%, sensitivity decreased to 16% (95%CI: 1–63), but specificity remained at 99% (95%CI: 98–100). None of the cases with low 5-min Apgar score exhibited prolonged, severe or repetitive decelerations.

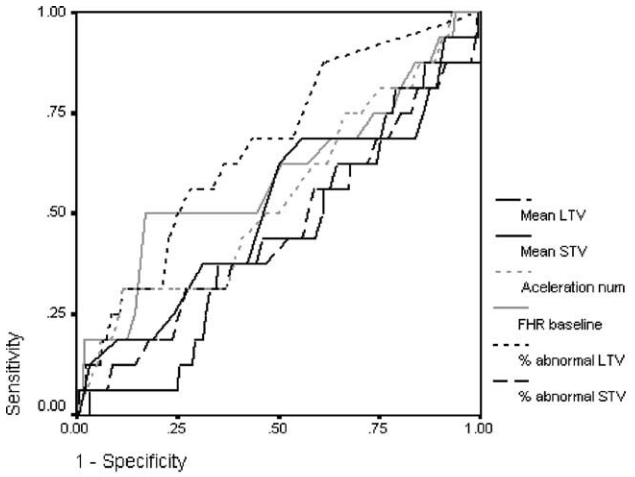
3.3. Prediction of umbilical artery acidemia

The ROC curves evaluating the discriminative capacity of FHR parameters to predict arterial pH <7.20 (46 cases) and <7.15 (15 cases) are displayed in Fig. 4. It should be noted that, of the eight cases with 1-min Apgar ≤4, only three had

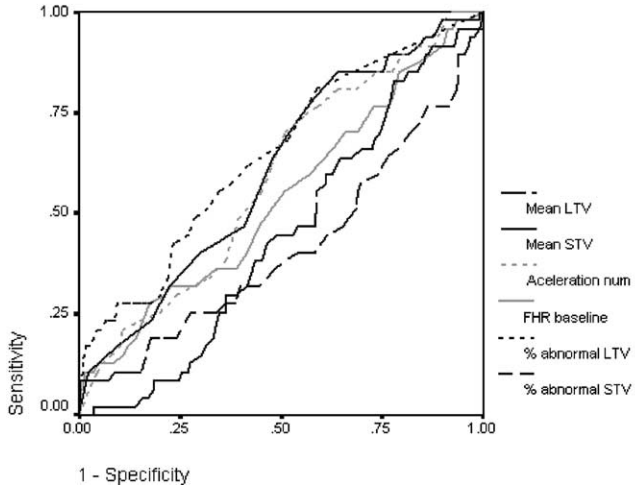


FHR parameters	Area	95% CI		Best
		Lower	Upper	
Acceleration num.	0.817	0.536	1.000	0.5
% abnormal STV	0.886	0.768	1.000	63.5
Mean STV	0.810	0.548	1.000	0.55
% abnormal LTV	0.854	0.614	1.000	30.5
Mean LTV	0.788	0.723	0.854	7.75
FHR baseline	0.637	0.348	0.927	143.5

Fig. 3. ROC curves for the prediction of 5-min Apgar ≤6 ($n = 6$) by the studied FHR parameters. On the right, areas under the ROC curves, 95% confidence intervals, and best cut-off values for each parameter. STV: short-term variability, LTV: long-term variability.



FHR parameters	Area	95% CI	
		Lower	Upper
Acceleration num.	0.634	0.467	0.800
% abnormal STV	0.508	0.319	0.697
Mean STV	0.545	0.344	0.746
% abnormal LTV	0.693	0.543	0.843
Mean LTV	0.451	0.279	0.623
FHR baseline	0.602	0.410	0.794



FHR parameters	Area	Ass. 95% CI	
		Lower	Upper
Acceleration num.	0.605	0.520	0.690
% abnormal STV	0.426	0.326	0.525
Mean STV	0.622	0.539	0.707
% abnormal LTV	0.662	0.576	0.749
Mean LTV	0.446	0.364	0.529
FHR baseline	0.524	0.431	0.617

Fig. 4. ROC curves and areas under the curve for prediction of arterial pH <7.15 (top, $n = 15$) and pH <7.20 (bottom, $n = 46$). STV: short-term variability, LTV: long-term variability).

umbilical blood-gas parameters that were considered valid (pH values 7.13, 7.16 and 7.19).

Absence of accelerations combined with abnormal long-term variability >30.5% obtained a sensitivity of 11% (95%CI: 4–24) and a specificity of 99% (95%CI: 97–100) in prediction of arterial pH <7.20. For arterial pH <7.15 these values were 12% (95%CI: 2–39) and 98% (95%CI: 96–99), respectively. Combining absence of accelerations with abnormal short-term variability >63.5%, provided a sensitivity of 11% (95%CI: 4–24) and a specificity of 98% (95%CI: 95–99) in prediction of arterial pH <7.20, and a sensitivity of 12% (95%CI: 2–39) and a specificity of 97% (95%CI: 94–99) in prediction of arterial pH<7.15. When abnormal long-term variability >30.5% was combined with abnormal short-term variability >63.5%, the sensitivity in prediction of arterial pH <7.20 was 11% (95%CI: 4–24) and the specificity 99% (95%CI: 97–100). For arterial pH <7.15 these values were 12% (95%CI: 2–39) and 98% (95%CI: 96–99). The combination of all three parameters provided a sensitivity of 11% (95%CI: 4–24) and a specificity of 99% (95%CI: 97–100) in predicting arterial pH <7.20. For arterial pH <7.15 these values were 12% (95%CI: 2–39) and 98% (95%CI: 96–99).

4. Comment

Our results show that SisPorto®2.0 FHR parameters related to the quantification of accelerations and variability have an excellent discriminative capacity to predict 1 and 5-min Apgar scores, and to a much lesser degree umbilical artery acidemia. The best cut-off values determined from analysis of accelerations, short- and long-term variability detected both cases that went on to develop hypoxic-ischemic encephalopathy. Addition of FHR baseline >150 bpm to the quantification of accelerations and variability decreased the sensitivity for prediction of Apgar scores significantly, while specificity remained high. None of the cases with low Apgar scores exhibited prolonged, severe or repetitive decelerations.

The programs' classification of decelerations evolved form the more traditional categorization as early, variable and late. In previous versions of the program when the latter was employed, a comparison between computer and experts regarding the identification of decelerations, fell short due to the surprisingly poor interobserver agreement [16] and only marginal consensus found with visual analysis [17]. This led us to question the clinical applicability of this classification [17], in spite of the well documented pathophysiological mechanisms underlying it, as demonstrated in controlled laboratory animal studies. It is possible that the different pathophysiological mechanisms involved in the etiology of decelerations usually occur simultaneously during labor, and the complexity of the originated FHR signal does not allow a consistent categorization into one of the predefined patterns. The simplified classification used in the current version of the program was derived from an optimisation study involving a limited number of patients [18], the validity of which can only truly be evaluated in a study carried out in the intrapartum.

The closer relationship between FHR parameters and Apgar score, as opposed to umbilical artery pH, seems clearly justified from a physiological point of view. Being mainly an evaluator of neurological and cardiac function, it follows that FHR monitoring should be a better predictor of neonatal parameters that are closely related (reflexes, muscle tone, and heart rate) rather than circulatory acid–base balance. The susceptibility of neurological and cardiac function to acidemia is likely to be related to its duration and swiftness of installation, as well as to individual cellular compensation mechanisms.

Although the best areas under the ROC curve were obtained with analysis of long-term variability, it was curious to find that quantification of short-term variability was also a good predictor of Apgar scores, in spite of the autocorrelation techniques routinely employed to improve external FHR signal quality. It has been well-demonstrated that this methodology provides only an approximation of the true beat-to-beat variability, and a value which is, on average, lower than that obtained by analysis of electrocardiographic signals [19,20].

While in the primordials of FHR monitoring it was established that its main objective was the prevention of fetal death and neurological handicap, it has since been demonstrated that these situations are frequently associated with fetal malformations, infections, metabolic diseases, and hypoxia occurring in the post-natal period [21]. It is not reasonable to expect FHR monitoring to identify or prevent such conditions, and so neonatal indicators of intrapartum hypoxia remain the only reasonable standards with which to compare the technique. The value of the Apgar score as an indicator of neonatal outcome has been questioned, because the vast majority of newborns with low estimates are known to have a normal post-natal course, it is subject to some interobserver variation [22] and can be affected by non-hypoxic conditions such as prematurity, drugs, pre-existing neurological conditions, resuscitation methods, and birth trauma [23]. On the other had, it has a higher association with neonatal death and long-term neurological handicap than umbilical artery pH [24,25] and is significantly associated with neonatal hypoxic-ischemic encephalopathy, a condition highly predictive of long-term neurological sequelae [26]. It is almost certain that in the group of newborns with low Apgar scores and/or low umbilical artery pH are included all of those who will die or develop severe handicaps as a consequence of peri-partum fetal hypoxia.

It is possible that the high discriminative capacity to predict Apgar scores was related to the inclusion of very premature fetuses, where fewer accelerations, decreased variability, as well as lower Apgar scores usually occur [24,27], and that there was some interdependency between these parameters. However, only half of the cases with 1-min Apgar $\leq$4 had a gestational age of less than 32 weeks, and these were the only cases in this premature group with no accelerations and abnormal variability above the chosen cut-off values. It thus seems likely that these FHR parameters maintain their discriminative capacity in very premature fetuses. Several other authors have found no significant differences in computer quantification of variability along the third trimester [28,29]. In particular, the lower limit of normality seems to remain relatively constant throughout this period [28,30,31].

The absence of fetuses with metabolic acidosis in this antepartum population delivered by cesarean section, in spite of documented abnormal neonatal outcomes presumably due to intra-uterine hypoxia, suggests that the cut-off values used to establish this diagnosis, derived predominantly from the study of intrapartum cases, may not be adequate in this scenario. Further research is needed to clarify this issue.

The Sonicaid® System 8000/8002 has been extensively evaluated in the antepartum, generally by analyzing the correlation between FHR parameters and neonatal outcome indicators, rather than their discriminative capacity [30,32,33]. Studies comparing FHR parameters with subsequent umbilical venous blood gas analysis obtained by cordocentesis found a high correlation between long-term variation and fetal hypoxemia [34], but the sensitivity of abnormal tracings in prediction of moderate to severe hypoxemia was only 45% [35].

Guzman et al. [36] studied 38 fetuses with intrauterine growth restriction and changes in uterine and/or umbilical Doppler waveforms, evaluating 1-h tracings acquired by System 8000 in the 4 h preceding cesarean birth. In the prediction of umbilical artery pH <7.20 ($n = 8$), short-term variation provided an area below the ROC curve of 0.97 (standard error 0.05). Using the best cut-off values determined by ROC curve analysis, short-term variation and absence of accelerations obtained sensitivities of 100%, and specificities of 80 and 73%, respectively, in the prediction of this outcome. Confidence intervals for these results were not reported.

Anceschi et al. [37] retrospectively evaluated 70 uncomplicated term pregnancies, where antepartum tracings lasting more than 40 min were acquired by System 8002 in the 4 h preceding elective cesarean section. No significant correlations between FHR parameters and umbilical artery blood gas values were found. Quantification of accelerations provided a sensitivity of 43% and a specificity of 28% in prediction of a good neonatal outcome, defined as arterial pH >7.20, extracellular fluid base deficit <4 mmol/l and 5-min Apgar >7. For fetal movements sensitivity was 67% and specificity 57%.

Using the 2CTG® system, Arduini et al. [38] studied 37 fetuses with intrauterine growth restriction and altered umbilical diastolic flow, evaluating tracings acquired immediately before elective cesarean section. Average long-term variability obtained a sensitivity of 77% and a specificity of 90% in prediction of umbilical artery pH <7.21. For average short-term variability these values were 83 and 100%, respectively.

The apparently contradictory results found in these publications are probably due to differences in the selected study population. In the relatively small number of cases studied with intra-uterine growth restriction and umbilical or uterine flow changes, a reasonable capacity of FHR variability to predict acidemia is apparent [36,38], but this is not the case in healthy term pregnancies [37]. It is evident that FHR monitoring needs to be a good discriminator in both of these populations, but independently of such considerations, the relatively small number of cases evaluated in these studies casts some uncertainties on the interpretation of results.

Our results suggest that automated quantification of accelerations and variability is a useful adjunct to the interpretation of antepartum FHR tracings, as it provides a good prediction of Apgar scores and, to a much lesser degree umbilical artery acidemia. This advantage could be related mainly to the quantification of long- and short-term variability, parameters in which eyeball analysis has a more evident limitation. Considering all studies that aimed to evaluate computerised antepartum FHR analysis and took care to avoid biases, the number of cases evaluated in this study is unprecedented, but in spite of this, the relatively few poor fetal outcomes precluded a more confident evaluation of the method's sensitivity, a fact that is not translated by overall ROC curve confidence intervals. The low incidence of newborns with poor neonatal outcome is a current reality in industrialized countries [39]. Further studies are therefore needed to confirm the high sensitivity obtained in prediction of Apgar scores.

Acknowledgements

The study was financed by scientific grants from the Portuguese government: JNICT PECS/C/SAU/207/95 and Praxis XXI 2/2.1/SAU/1351/95.

References

[1] Devoe LD, Castillo RA, Sherline DM. The nonstress test as a diagnostic test: a critical reappraisal. Am J Obstet Gynecol 1985;152:1047–53.
[2] Thacker SB, Berkelman RL. Assessing the diagnostic accuracy and efficacy of selected antepartum fetal surveillance techniques. Obstet Gynecol Surv 1986;41:121–41.
[3] Mohide P, Keirse JNC. Biophysical assessment of fetal well-being. In: Chalmers I, Enkin M, Keirse MJNC, editors. Effective care in pregnancy and childbirth. Oxford: Oxford University Press; 1989. p. 477–92.
[4] Paneth N, Bommarito M, Stricker J. Electronic fetal monitoring and later outcome. Clin Invest Med 1993;16:159–65.
[5] Pattison N, McCowan L. Cardiotocography for antepartum fetal assessment. Cochrane Database Sys Rev 2000;(2):CD001068.
[6] Grant JM. Clinical overview of cardiotocography. Br J Obstet Gynaecol 1993;100(Suppl 9):4–7.
[7] Dawes GS, Redman CW, Smith JH. Improvements in the registration and analysis of fetal heart rate records at the bedside. Br J Obstet Gynaecol 1985;92(4):317–25.
[8] Ayres-de-Campos D, Bernardes J, Garrido A, Marques-de-Sá J, Pereira-Leite L. SisPorto 2.0: a program for automated analysis of cardiotocograms. J Matern Fetal Med 2000;9(5):311–8.
[9] Hollmen AI, Jouppila R, Koivisto M, Maatta L, Pihlajaniemi R, Puukka M, et al. Neurologic activity of infants following anesthesia for cesarean section. Anesthesiology 1978;48(5):350–6.
[10] Downing JW, Houlton PC, Barclay A. Extradural analgesia for caesarean section: a comparison with general anaesthesia. Br J Anaesth 1979;51(4):367–74.
[11] Crawford JS, James FM, Davies P, Crawley M. A further study of general anaesthesia for caesarean section. Br J Anaesth 1976;48(7):661–7.
[12] Huch A, Huch C, Rooth G. Guidelines for blood sampling and measurement of pH and blood gas values in Obstetrics. Eur J Obstet Gynecol Reprod Biol 1994;54:165–75.
[13] ACOG technical bulletin. Umbilical artery blood acid-base analysis. Int J Gynecol Obstet 1996;52:305–10.
[14] Westgate J, Garibaldi JM, Greene KR. Umbilical cord blood gas analysis at delivery: a time for quality data. Br J Obstet Gynaecol 1994;101(12):1054–63.
[15] Altman DG, Bland JM. Statistics notes: diagnostic tests 3: receiver operating characteristic plots. BMJ 1994;309(6948):188.
[16] Bernardes J, Costa-Pereira A, Ayres-de-Campos D, van Geijn HP, Pereira-Leite L. Evaluation of interobserver agreement of cardiotocograms. Int J Gynecol Obstet 1997;57:33–7.
[17] Ayres-de-Campos D, Bernardes JE. Variable and late decelerations: can a consensus be reached in their identification. Int J Gynecol Obstet 1999;65:305–6.
[18] Bernardes J, Ayres-de-Campos D, Costa-Pereira A, Pereira-Leite L, Garrido A. Objective computerized fetal heart rate analysis. Int J Gynecol Obstet 1998;62:141–7.
[19] Spencer JAD, Belcher R, Dawes GS. The influence of signal loss on the comparison between computer analyses of the fetal heart rate in labour using pulsed Doppler ultrasound (with autocorrelation) and simultaneous scalp electrocardiogram. Eur J Obstet Gynecol Reprod Biol 1987;25(1):29–34.
[20] Lawson GW, Belcher R, Dawes GS, Redman CWG. A comparison of ultrasound (with autocorrelation) and direct electrocardiogram fetal heart rate detector systems. Am J Obstet Gynecol 1983;147(6):721–2.
[21] ACOG technical bulletin. Fetal and neonatal neurologic injury. Int. J. Gynecol. Obstet. 41 (1993) 97–101.
[22] Lafeber HN, van Geijn HP. Endpoints of fetal condition and neonatal assessment. In: van Geijn HP, Copray FJA, editor. A critical appraisal of fetal surveillance, vol. 9. Amsterdam: Elsevier; 1994. p. 70–3.

[23] Hankins GV. Apgar scores: are they enough. Contemporary Ob/Gyn 1991;36:13–26.
[24] Casey BM, McIntire DD, Leveno KJ. The continuing value of the Apgar score for the asessment of newborn infacts. N Engl J Med 2001;344:467–71.
[25] Dennis J, Johnson A, Mutch L, Yudkin P, Johnson P. Acid-base status at birth and neurodevelopmental outcome at four and one-half years. Am J Obstet Gynecol 1989;161(1):213–20.
[26] van de Riet JE, Vandenbussche FPHA, Le Cessie S, Keirse MJNC. Newborn assessment and long-term adverse outcome: a systematic review. Am J Obstet Gynecol 1999;180(4):1024–9.
[27] Mantel R, Van Geijn HP, Ververs IAP, Colenbrander GJ, Kostense PJ. Automated analysis of antepartum fetal heart rate in relation to fetal rest-activity states: a longitudinal study of uncomplicated pregnancies using the Sonicaid System 8000. Eur J Obstet Gynecol Reprod Biol 1996;71:41–51.
[28] Dawes GS. Computerised analysis of the fetal heart rate. Eur J Obstet Gynecol Reprod Biol 1991;42(Suppl):S5–8.
[29] Arduini D, Rizzo G, Piana G, Bonalumi A, Brambilla P, Romanini C. Computerized analysis of fetal heart rate: I. Desciption of the system (2CTG). J Matern Fetal Invest 1993;3: 159–63.
[30] Dawes GS, Moulden M, Redman CWG. Short-term fetal heart rate variation, decelerations, and umbilical flow velocity waveforms before labor. Obstet Gynecol 1992;80(4):673–8.
[31] Snijders RJM, McLaren R, Nicolaides KH. Computer-assisted analysis of fetal heart rate patterns at 20–41 weeks gestation. Fetal Diagn Ther 1990;5(2):79–83.
[32] Henson GL, Dawes GS, Redman CWG. Antenatal fetal heart-rate variability in relation to fetal acid-base statuse at casarean section. Br J Obstet Gynaecol 1983;90(6):516–21.
[33] Smith JH, Anand KJ, Cotes PM, Dawes GS, Harkness RA, Howlett TA, Rees LH, Redman CWG. Antenatal fetal heart rate variation in relation to the respiratory and metabolic status of the compromised human fetus. Br J Obstet Gynaecol 1988;95(10):980–9.
[34] Visser GHA, Sadovsky G, Nicolaides KH. Antepartum heart rate patterns in small-for-gestational-age third-trimester fetuses: correlations with blood gas values obtained at cordocentesis. Am J Obstet Gynecol 1990;162(3):698–703.
[35] Nicolaides KH, Sadovsky G, Visser GH. Heart rate patterns in normoxemic, hypoxemic, and anemic second-trimester fetuses. Am J Obstet Gynecol 1989;160(5 Pt 1):1034–7.
[36] Guzman ER, Vintzileos AM, Martins M, Benito C, Houlihan C, Hanley ME. The efficacy of individual computer heart rate indices in detecting acidemia at birth in growth-restricted fetuses. Obstet Gynecol 1996;87:969–74.
[37] Anceschi MM, Piazze JJ, Vozzi G, Ruozi Berretta A, Figliolini C, Vigna R, Cosmi EV. Antepartum computerized CTG and neonatal acid-base status at birth. Int J Gynaecol Obstet 1999;65(3):267–72.
[38] Arduini D, Rizzo G, Romanini C. Computerized analysis of fetal heart rate in normal and growth-retarded fetuses. In: Kurjac A, Chervenak FA, editors. The fetus as a patient. London: Parthenon Publishing Group; 1994. p. 289–97.
[39] Thorngren-Jerneck K. Herbst A.Low 5-minute Apgar score: a population-based register study of 1 million term births. Obstet Gynecol 2001;98(1):65–70.