

Effect of signal acquisition method on the fetal heart rate analysis with phase rectified signal averaging

J.A.M. van Scheepen, MSc^{1,2}, M.P.H. Koster^{2,3} MD, PhD. , B. Vasak MD², em. prof. C. Redman⁴ and Prof. A. Franx² MD PhD, A. Georgieva, PhD^{4,5}

¹ University of Twente, Technical Medicine

² Division Woman and Baby, University Medical Centre, Utrecht, the Netherlands

³ Department of Obstetrics and Gynaecology, Erasmus University Medical Center, Rotterdam, the Netherlands

⁴ Oxford Centre for Fetal Monitoring Technologies, Nuffield Department of Obstetrics and Gynaecology, University of Oxford, United Kingdom

Acknowledgements: A Georgieva is funded by the Action Medical Research and the Henry Smith Charity.

Corresponding Author: J.A.M. van Scheepen, j.a.m.vanscheepen@outlook.com

Abstract

Phase rectified signal averaging (PRSA) is increasingly used for fetal heart rate monitoring, both with traces acquired with external Doppler cardiotocography (D-FHR), and with transabdominal fetal electrocardiography (ta-FHR). However, it is unclear to what extent the acquisition method influences the PRSA analysis, whether results from using one acquisition method are comparable to those based on FHR acquired by the other method, and if not, which should be the preferred method.

To address these questions, we applied PRSA analysis to 28 antepartum synchronous recordings of the fetal heart rate using simultaneously D-FHR and ta-FHR. The data included late-onset intrauterine growth restricted (IUGR) fetuses ($n=20$) and non-IUGR fetuses ($n=8$), all of them at gestation ≥ 34 weeks. PRSA analysis depends on two parameters intrinsic to the algorithm, T and S. We analyzed the data using parameters that included all values adopted by other researchers previously (derived from a literature search in PubMed and Google Scholar). T and S were adjusted for the difference in acquisition techniques. We found that the correlation between PRSA analysis based on D-FHR and ta-FHR decreased with decreasing values of the PRSA parameters T and S. Therefore, the acquisition technique affects PRSA values for high resolution PRSA (low values of T and S).

In conclusion, for low resolution PRSA, the results from both acquisition methods are comparable. Because ta-FHR signals provide beat to beat data and thus capture more subtle differences in the heart rate variation than D-FHR signals (pre-processed by commercial monitors), we assumed that ta-FHR may provide potentially valuable extra information compared to D-FHR. However, no parameter settings or acquisition method seemed to have diagnostic value for identifying the late-onset IUGR babies in our dataset.

Introduction

A major challenge in obstetric care is to identify fetuses that may benefit from earlier or immediate delivery. Continuous fetal heart rate monitoring is used to inform this decision and comprises two time series graphs: fetal heart rate (FHR) and uterine contractions. This is called cardiotocography (CTG (Alfirevic et al. 2013; Graatsma et al. 2012)). Usually, the CTG is interpreted visually by clinicians – a process that is highly susceptible to inter- and intra-observer variability (Alfirevic et al. 2013). Computerized methods for CTG analysis aim to resolve this problem by providing objective standardized measures of deviations from normality. A newly-emerging computer-based algorithm for FHR analysis, called phase rectified signal averaging (PRSA), has been under investigation (Graatsma et al. 2012; Lobmaier et al. 2012). The FHR can be analyzed with PRSA both prior to and during labour to provide objective decision support for the identification of the fetus at risk of compromise (Huhn & Lobmaier 2011; Graatsma et al. 2012; Fanelli et al. 2014; Fanelli et al. 2013; Signorini et al. 2014; Georgieva et al. 2014; Stampalija et al. 2015), including intrauterine growth restriction (IUGR) and acidosis. However, reports on antenatal PRSA have used FHR traces acquired by two fundamentally different devices: the standard external Doppler ultrasound (D-FHR) and the newer transabdominal fetal ECG monitors (ta-FHR) (Graatsma et al. 2009). Doppler ultrasound uses the reflection of a sound wave from the moving fetal heart, an autocorrelation function and zero-order interpolation to calculate an *average* heart rate in beats per minute. The device outputs the equally sampled FHR at 4Hz. The transabdominal monitor measures the entire electrical field on the maternal abdomen with significantly higher sample frequency (1000 Hz (Graatsma et al. 2012)). Then it uses signal processing to separate the raw fetal ECG signal from the maternal ECG and other electrical activity. Thus, this device provides the *true beat-to-beat* heart rate in beats per minute which is sampled unevenly (Huhn & Lobmaier 2011; Graatsma et al. 2012; Lobmaier et al. 2012; Abdulhay et al. 2014; Graatsma et al. 2009; Cesarelli et al. 2007; Hasan et al. 2009; Piéri et al. 2001). The AN24 monitor (Monica Healthcare, Nottingham), used here and in several of the PRSA studies in the literature, has been shown to produce CTG that is similar to standard Doppler CTG (Cohen et al. 2012;

Reinhard et al. 2012; Sanger et al. 2012). Ta-FHR has been reported to be more closely related to internal scalp electrode than is D-FHR (Cohen et al. 2012), but there is generally limited evidence. Ta-FHR has been reported to have higher signal loss in general during the second stage of labour (personal communication, Monica Healthcare), but it has reportedly better signal quality in women with higher BMI (Graatsma et al. 2010). The signals obtained with internal scalp electrode are widely considered as a ‘reference signal’ for FHR monitoring intrapartum. However, to our knowledge, PRSA analysis has not been applied to FHR recorded with this method specifically and has been usually applied to antepartum recordings (prior to rupture of the membranes). Therefore, our study focused on the use of external FHR monitoring only.

The impact on FHR monitoring of differences in acquisition modes on computerized analysis has been investigated (Gonalves et al. 2015; Gonalves et al. 2006). However, it has not been previously reported whether computerized analysis of ta-FHR is comparable to that of standard D-FHR. While PRSA analysis seems promising with either acquisition method, it is not clear if the conclusions drawn using data from one device are transferable to the other. Moreover, true beat-to-beat FHR (ta-FHR) may, in principle, hold a higher clinical significance because it does not use averaging and is thus more precise. This raises the question whether the acquisition method affects the PRSA calculation and leads to our hypothesis that different methods of signal acquisition may impose important differences in computerized measures such as PRSA.

Methods

Design

This was a retrospective case-control study (Figure 1) comparing PRSA analysis of D-FHR and ta-FHR. PRSA is typically characterized by two components – the average acceleration capacity (AAC) and average deceleration capacity (ADC) (Huhn & Lobmaier 2011; Georgieva et al. 2014). The aim of this study was to determine the correlation between PRSA values derived from the two acquisition techniques. FHR traces were acquired prior to labour using both standard Doppler and transabdominal monitoring. We compared the PRSA values when calculated using D-FHR and ta-FHR for the full range

of possible algorithm parameters, T and S. The comparison was conducted using the Pearson correlation coefficient and by assessing the agreement of the methods with Bland-Altman plots (Bland & Altman 1986). For the correlation coefficient p-values for testing the hypothesis of no correlation are also calculated using corrcoeff (Matlab). Mean and relative absolute difference between both acquisition methods's AAC and ADC were also reported. Finally, we compared the PRSA analysis (both ta-FHR and D-FHR) for IUGR and control babies. The literature on PRSA applied to fetal heart rate monitoring was reviewed and reported. A search was performed in Pubmed and Google scholar with the search words: 'fetal phase rectified signal averaging'.

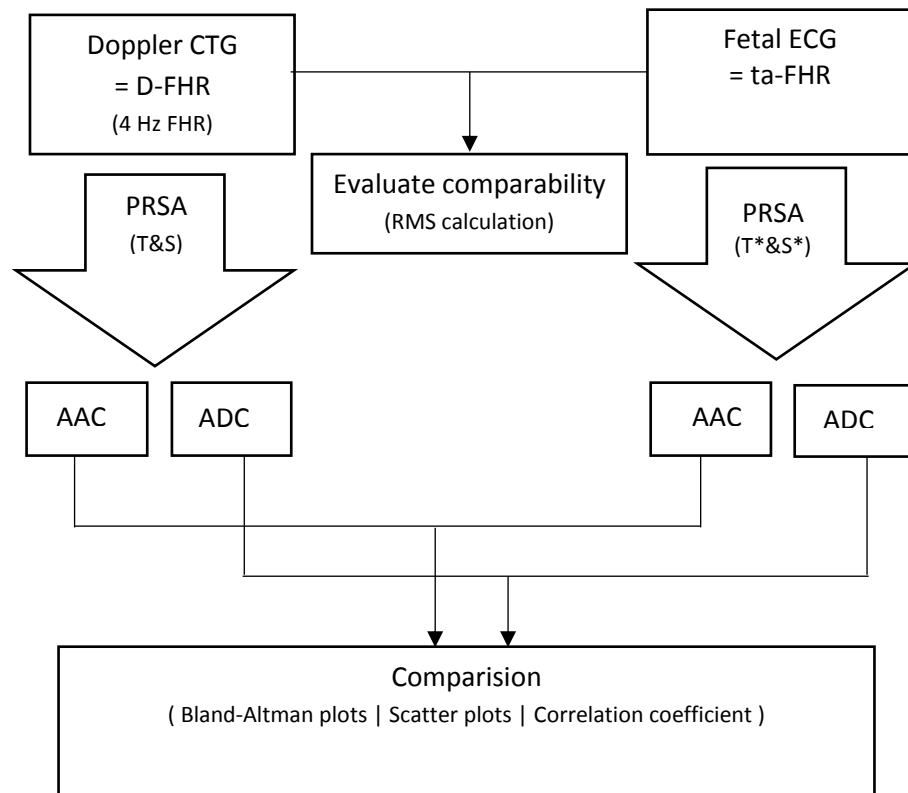


Figure 1: Study design: method comparison. RMS – root mean squared error. T and S are adjusted to the difference in sample frequency between ta-FHR and D-FHR. Therefore T & S denote the parameter settings applied to D-FHR; T* & S* denote the parameter settings applied to ta-FHR.

Data

We studied 28 synchronous D-FHR and ta-FHR recordings, 20 from 9 IUGR fetuses and 8 from 4 non-IUGR fetuses (multiple records per person at different gestational ages were allowed). The diagnosis of IUGR was based on standard clinical practice which included Doppler blood flow pulsatility and ultrasound biometry. The IUGR group consisted of fetuses that showed growth retardation with an

onset only after 34 weeks of gestation. The control group consisted of fetuses that were monitored for standard care. The study was approved by the institutional Medical Ethical Committee of the University Medical Center Utrecht. Written informed consent was obtained from the participants. Only singleton pregnancies were included and fetuses with known congenital abnormalities were excluded.

The Doppler CTG recordings were acquired using an external CTG monitor (Philips Healthcare, The Netherlands) accordance to standard care in the Wilhelmina Children's Hospital in Utrecht. These were women visiting the clinic for routine check up of pregnancy or an extra check up, for example in case of reduced fetal movements. The fetal ECG signals were simultaneously collected using the Monica AN24 system (Monica Healthcare, Nottingham). This was a supplementary recording after informed consent. Doppler CTG has a sampling rate of 4 Hz and is indicated in beats per minute. From the Monica AN24 monitor, we used the true beat-to-beat RR interval in millisecond (ms). This was then rescaled to FHR in beats per minute (bpm) by dividing 60000 by the RR intervals which is not uniformly sampled. To assure that the same time window of FHR monitoring is used for the comparison, the time stamp was corrected if needed. A time correction was applied to the time stamp by calculation of the cross correlation.

Currently, for ta-FHR the Monica is the only commercial manufacturer. For Doppler FHR there are various manufacturers. There might be minor differences between FHR with different monitors resulting from rounding errors. However they use the same principle (e.g. Doppler), it is therefore expected that the results would be similar with equipments from other manufacturers.

Analysis

The PRSA algorithm was constructed according to the original algorithm of Bauer et al. (Bauer et al. 2006), adapted for FHR analysis by Huhn et al. (Huhn & Lobmaier 2011). Decelerations and accelerations of the FHR were *not* excluded from the calculation. Usually, the PRSA algorithm uses three parameters: T, L and S. The parameter T determines how many data points are taken into

account for the anchor point selection, the parameter L determines how many data points are used to calculate the PRSA series (or the time window analyzed) and the parameter S determines how many data points are used to calculate the average accelerative capacity (AAC) and the average decelerative capacity (ADC). We used AAC and ADC only, and thus the parameter L was redundant because only the AAC and ADC were considered and not the entire phase rectified curve (only T and S will be reported below). Majority of research has focused on AAC and ADC, while one research group works with the Acceleration Phase Rectified slope (Fanelli et al. 2013; Fanelli et al. 2014; Signorini et al. 2014; Tagliaferri et al. 2015). Similarly to AAC and ADC, it characterizes the steepness of the slope of the PRSA curve. For consistency, in this study we focused on the AAC and ADC.

The PRSA algorithm is a sample based algorithm (Bauer et al. 2006), thus it is not correct to use the same T and S values when analyzing the 4Hz D-FHR and the unevenly sampled ta-FHR, because it would result in analyzing over different time windows. Hence, we denote with T and S the parameters for the D-FHR signals (units 0.25 seconds) and with T* and S* as the corresponding parameters for the ta-FHR (units beats). T and S are linked to T* and S* through the following formula $T = \left(\frac{T^*}{\frac{144}{60}} \right) \times 4$.

This formula assumes an average fetal heart rate of 144 beats per minute and if needed the values were rounded. For consistency, it was important to use a constant average fetal heart rate in this conversion from 0.25s into beats (i.e. from T and S into T* and S*). The above formula is a simplification as the average fetal heart rate decreases with advancing gestation and varies between individuals. Therefore other values, including the extremes of 100bpm and 180bpm, were tested for the average fetal heart rate and did not change substantially the results of this study, i.e. the correlation of PRSA values based on D-FHR and ta-FHR (data not shown).

We used the Root Mean Squared (RMS) error to measure the differences between the D-FHR and corresponding ta-FHR. For each recording pair, one RMS error was calculated. Because the D-FHR is uniformly sampled at 4Hz and the ta-FHR is not, only the data points in the D-FHR (time-wise) closest to a data point in the ta-FHR were used for the RMS error calculation.

The correlation coefficients of AAC/ADC from PRSA analysis on D-FHR and AAC/ADC from PRSA analysis on ta-FHR are reported. T and S values of 1 to 5 (increments of 1); 5 to 10 (increments of 2); and 10 to 50 (increments of 5) were used together with their corresponding T* and S*. The results for two settings of the parameters are shown in scatter and Bland-Altman plots (Bland & Altman 1986). Analysis was conducted using MATLAB (version R2014b).

Results

The demographics of the dataset are shown in Table 1. The cases and controls had similar characteristics overall with IUGR babies having longer traces.

Table 1: Data demographics, reported is mean \pm standard deviation.

Data characteristics Population size: n=28	Control (n = 8)	IUGR (n =20)
Length registration (min)	36 \pm 4	41 \pm 18
Gestational age at recording (weeks+days)	34 \pm 4+0.6	36+3.6 \pm 1+2.8
BMI (2 missing)	25 \pm 1.7	23 \pm 2.6
Number of CTG records per subject	2 \pm 1.7	2.2 \pm 1.7
Birth weight (grams)	3093 \pm 290	2168 \pm 459
Difference between CTG recording and birth (days)	36 \pm 30	9.5 \pm 9

Figure 2 shows the RMS error comparing the pairs of D-FHR and ta-FHR. An empirically selected threshold at 14 bpm is used to divide the recording pairs in two groups of higher and lower RMS. These are used to color code the recording pairs in Figure 2, according to their RMS error (red and blue). Roughly two thirds of all recordings had a RMS error lower or equal to 14 bpm. Several other thresholds were used (20, 19, and 12 bpm) without significant change of the main results in this study. The high RMS error in pairs number 1 and 24 was due to long intervals caused by missing signals in the ta-FHR.

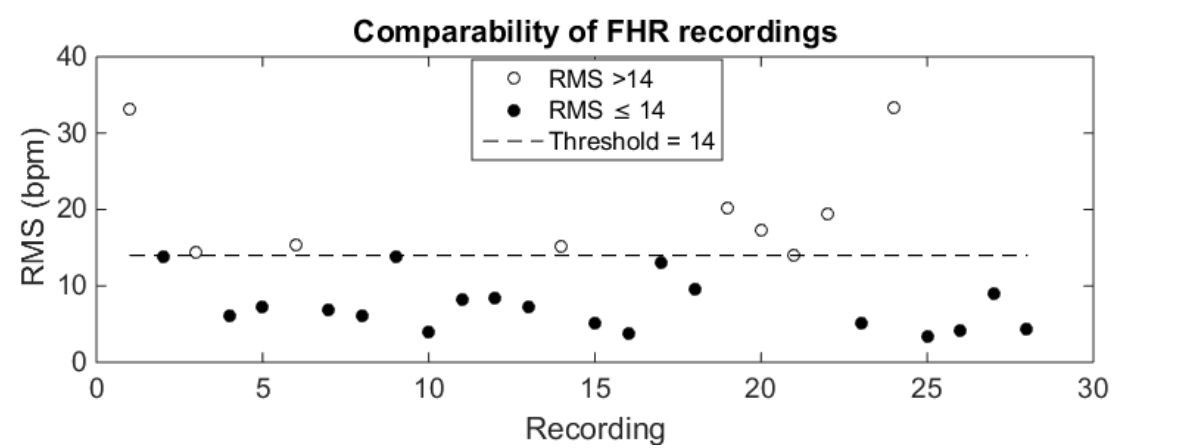


Figure 2: Root mean square error calculated to express comparability between registration pairs (D-FHR and ta-FHR). a threshold at 14 bpm divides the registration pairs in two categories (with a size of respectively 1/3 and 2/3 of all recordings).

The most common discrepancies we observed were more spiky/variable signals with ta-FHR and missing segments in ta-FHR. The faster oscillations (data not shown) were not present in the D-FHR, appearing to have been filtered out, which is not surprising given that the D-FHR signal is pre-processed with autocorrelation.

Nine applications of PRSA for fetal monitoring were found and listed in

Table 2. When more than one publication of the same PRSA settings was published by the same group, only the most recent publication was included in Table 2.

Table 2 Comprehensive list of the PRSA versions, described in literature (identified in the literature search). If not specified, the publication focused on antepartum registrations. T specifies the time window used for anchor point

selection; S specifies the time window used to calculate the average acceleration capacity and the average deceleration capacity.

Article	Detection method	T	S	Population size (IUGR/healthy)	Decelerations
Huhn et al. (2011)	D-FHR (4Hz)	10 s	10s	235(74/161)	excluded
Graatsma et al. (2012)	ta-FHR	According to Huhn et al.		120 (30/90)	excluded
Lobmaier et al. (2012)	D-FHR (4Hz)	10 s	T	82 (39/43)	excluded
Fanelli et al. (2013)	D-FHR (2Hz)	20 s	T	122 (61/61)	included
Fanelli et al. (2014)	D-FHR (2Hz)	20 s	T	120 (61/59)	Included
Signorini et al. (2014)	D-FHR	Not specified		122 (61/61)	Included
Rivolta et al. (2014)	ta-FHR	2-5 heart beats	T	9 sheep models with simulated labour-like insults	Included and excluded
Georgieva et al. (2014)	D-FHR (4Hz)	1.24s	11.25s	Cohort of 7568 labours, detecting acidosis at birth	Included
Stampalija et al. (2015)	ta-FHR	9 heart beats	T	59 (22/37)	Included

Figure 3 and Figure 4 show the correlation between AAC/ADC on ta-FHR and D-FHR across values for T and S ranging from 1 to 50 (0.25 s). There is good correlation (>0.8 with a p-value < 0.01) for average or high T and S values but the correlation is lower for T or S below 10 (0.25 s) and is down to 0.1 (with a p-value >0.05) for unit T or S. Figure 3 and Figure 4 show the correlations for the recording pairs with $RMS \leq 14$. The corresponding figures, when all recording pairs were included, had the same trend (data not shown). As the FHR traces included in Figures 3 and 4 were more comparable, we expect the results for the recording pairs shown in these Figures 3 and 4 to be more reliable.

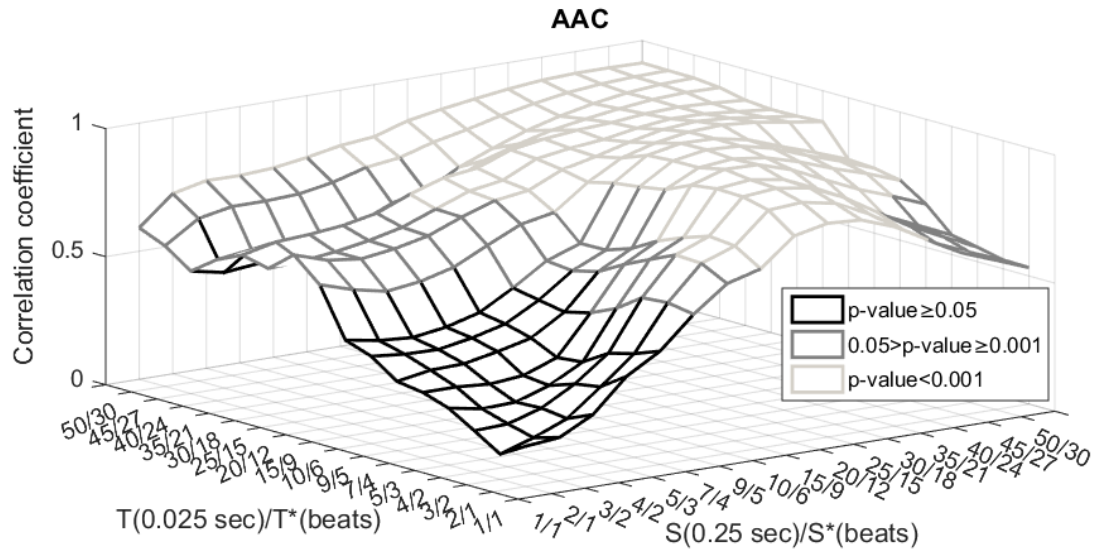


Figure 3: Pearson correlation for AAC calculated using D-FHR and ta-FHR for different settings for T and S and their corresponding T* and S* (data consists of the 19 pairs with RMS error<14bpm). The corresponding p-values are represented in color: black for $p \geq 0.05$, middle gray for $0.05 > p \geq 0.001$ and light gray for $p < 0.001$.

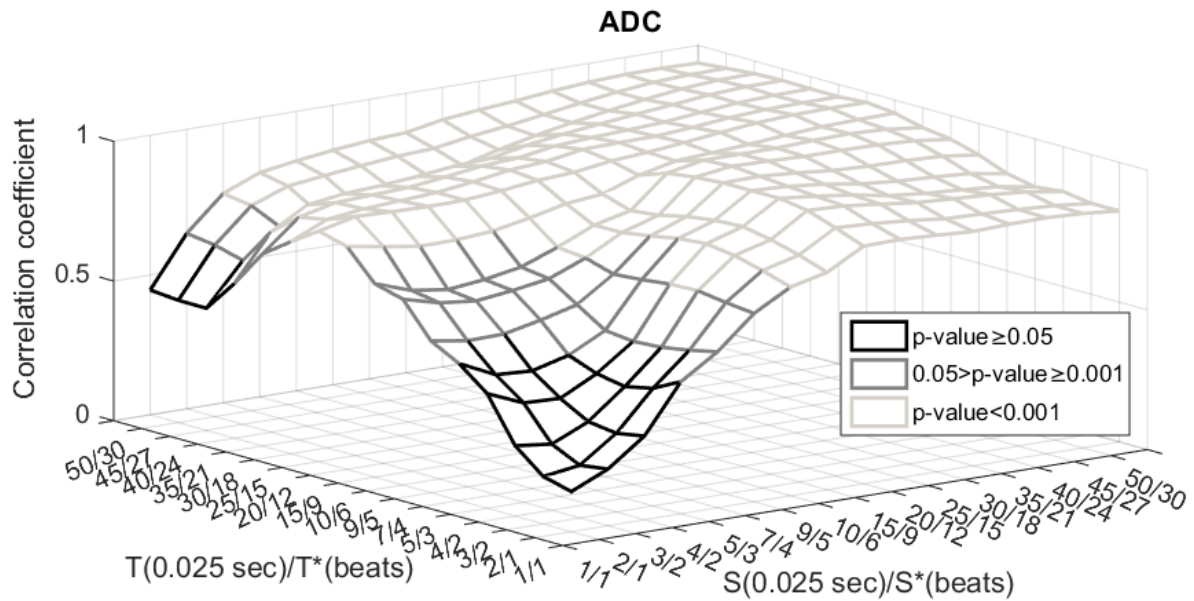


Figure 4 : Pearson correlation for ADC calculated using D-FHR and ta-FHR for different settings for T and S and their corresponding T* and S* (data consists of the 19 pairs with RMS error<14bpm). The corresponding p-values are represented in color: black for $p \geq 0.05$, middle gray for $0.05 > p \geq 0.001$ and light gray for $p < 0.001$.

Table 3 shows how the parameter settings used for PRSA analysis in previous studies performed in our analyses together with the corresponding correlation coefficients for the two groups (high and low RMS). As expected, the correlation between PRSA on ta-FHR and D-FHR is higher when only the pairs with low RMS are considered. For all settings of T and S in Table 3, there was overall good to high

correlation with coefficients typically between 0.68 and 0.9 when the RMS error is <14bpm. There was one exception and that are the settings proposed by Rivolta et al. which are also the smallest values of T and S. For higher values of T and S, the correlation increased. Two example settings from both extremes are considered in detail below.

Table 3: Correlation coefficient between PRSA analysis (on the total of 28 traces in our study) of D-FHR and ta-FHR for the parameter settings proposed and used in the literature (T and S are expressed in 0.25 seconds and T* and S* in beats). T and S are transformed to T* and S* to correct for the difference in acquisition technique.

Parameter settings according to:	T [0.25 s]	T* [beats]	S [0.25 s]	S* [beats]	Correlation coefficient AAC		Correlation coefficient: ADC	
					all (n≤28)	RMS≤14 (n≤19)	all (n≤28)	RMS≤14 (n≤19)
Huhn et al. (Huhn & Lobmaier 2011) Lobmaier et al. (Lobmaier et al. 2012) Graatsma et al. (Graatsma et al. 2012)	40	24	T	T*	0,50 p<0,01	0,85 p<0,01	0,72 p<0,01	0,90 p<0,01
Fanelli et al. (Fanelli et al. 2014; Fanelli et al. 2013)	20	48	T	T*	0,58 p<0,01	0,86 p<0,01	0,74 p<0,01	0,82 p<0,01
Rivolta et al. (Rivolta et al. 2014)	3	2	T	T*	0,14 p<0,5	0,17 p<0,48	0,06 p<0,8	0,18 p<0,45
Georgieva et al. (Georgieva et al. 2014)	5	3	45	27	0,77 p<0,01	0,68 p<0,01	0,80 p<0,01	0,81 p<0,01
Stampalija et al. (Stampalija et al. 2015)	15	9	T	T*	0,47 p<0,01	0,73 p<0,01	0,62 p<0,01	0,80 p<0,01

To investigate the PRSA discriminatory power for IUGR babies, Figure 5 and Figure 6 show the scatter plots for AAC and ADC based on D-FHR and ta-FHR on all of the 28 traces in our study. In Figure 5, the parameter settings proposed by Georgieva et al. ($T=1.25s$ and $S=11.25s$) were used, which had the highest correlation in Table 3. On the other hand, Figure 6 shows the scatter plots for the parameter settings proposed by Rivolta et al. ($T=S=3$ (0.25 s)), resulting in the lowest correlation in Table 3. Both figures illustrate that, regardless of the correlation of PRSA based on ta-FHR and D-FHR, there is no clear separation between IUGR and non-IUGR fetuses.

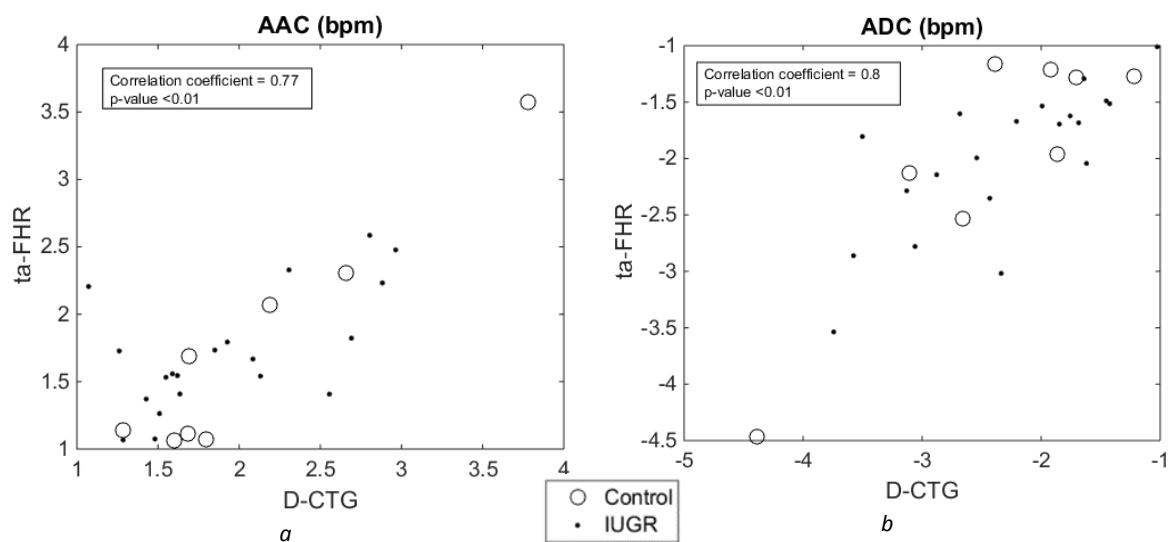


Figure 5: Scatter plots showing the relation between PRSA analysis on ta-FHR and D-FHR using the parameter values of Georgieva et al 2014. ($T=5$ (0.25s) and $S=45$ (0.25s)). The circles indicate registration pairs of control fetuses and the dots of IUGR fetuses.

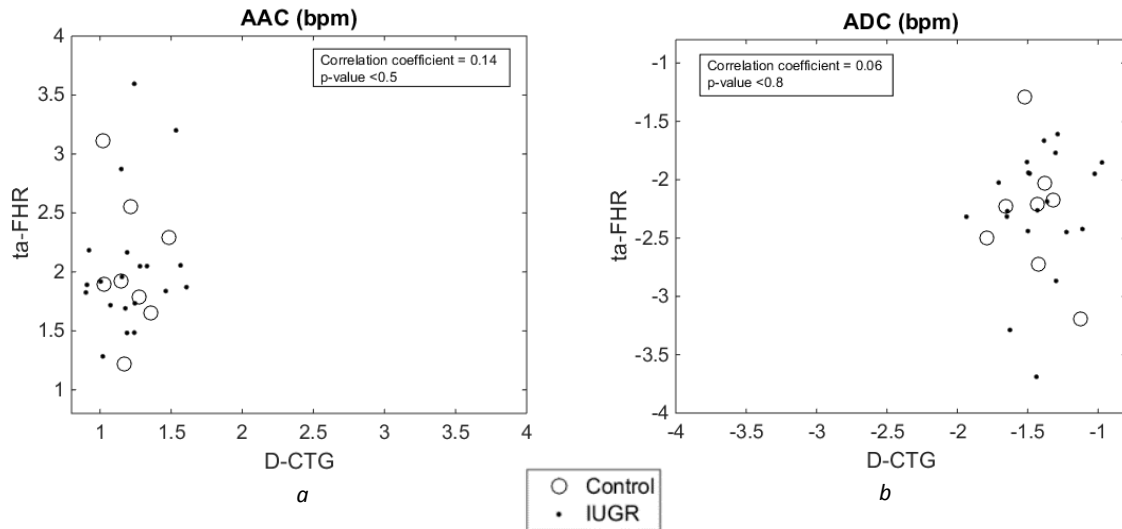


Figure 6: Scatter plots showing the relation between PRSA analysis on ta-FHR and D-FHR using the parameter values of Rivolta et al 2014. ($T=3(0.25s)$ and $S=3(0.25s)$). The circles indicate registration pairs of control fetuses and the dots of IUGR fetuses.

To investigate if there is a systematic error or proportional bias between PRSA derived from ta-FHR and D-FHR, the Bland-Altman plots corresponding to the scatter plots on Figure 5 and Figure 6 are shown in Figure 7 and Figure 8 respectively. There is evidence of a small systematic difference but most recordings fall within ± 1.96 standard deviations. There is no evidence of possible relationship of the differences between the PRSA on ta-FHR and D-FHR and their mean (i.e., proportional bias).

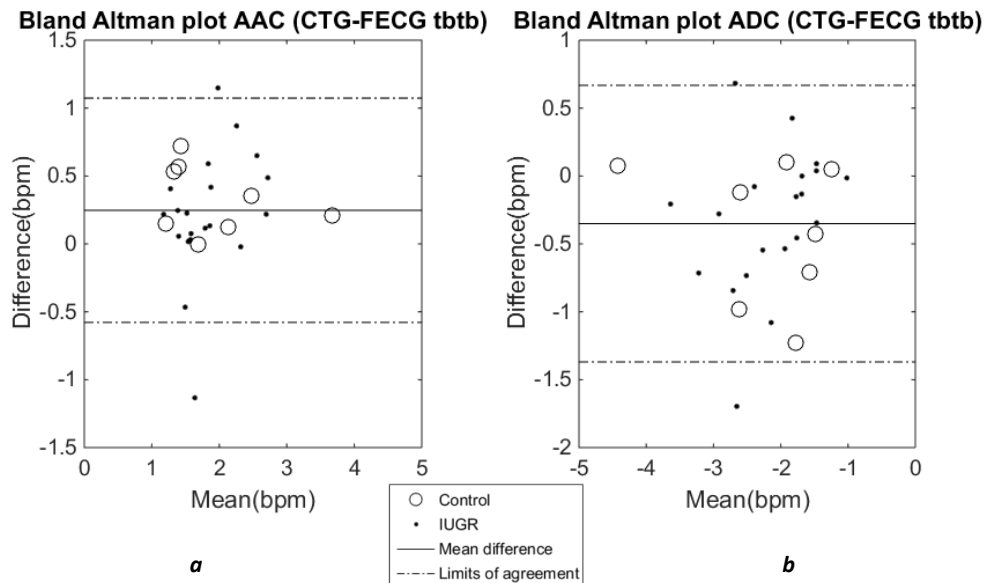


Figure 7: Bland-Altman plot investigating the relation between PRSA analysis on ta-FHR and D-FHR using the parameter values of Georgieva et al. ($T=5(0.25s)$ and $S=45(0.25s)$). The dash-dotted lines show the 95% limits of agreement. The circles indicate registration pairs of control fetuses and the dots of IUGR fetuses. The mean absolute difference was 0.36 for AAC and 0.46 for ADC. The mean relative difference was 0.23 for AAC and 0.26 for ADC.

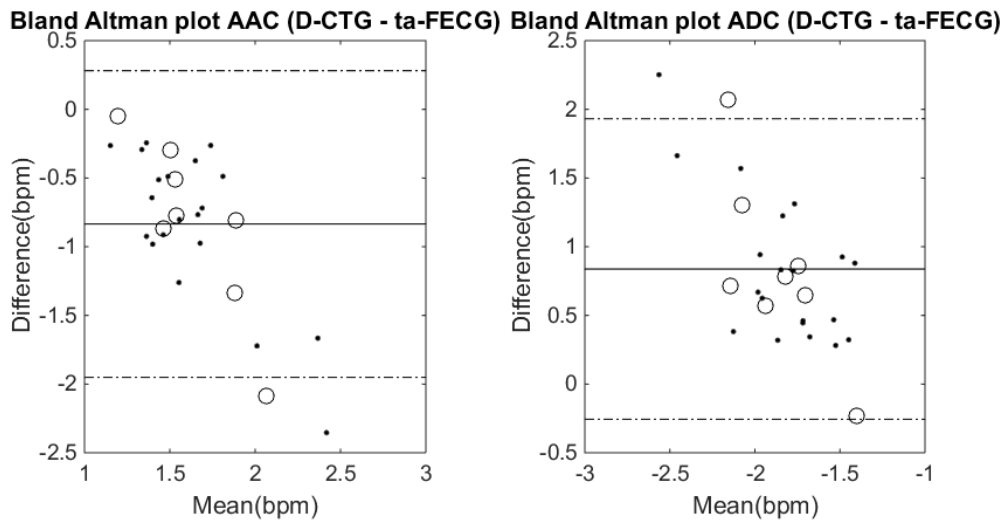


Figure 8: Bland-Altman plot investigating the relation between PRSA analysis on ta-FHR and D-FHR using the parameter values of Rivolta et al. ($T=S=3$ (0.25 s)). The dash-dotted lines show the 95% limits of agreement. The circles indicate registration pairs of control fetuses and the dots of IUGR fetuses. The mean absolute difference was 0.84 for AAC and 0.85 for ADC. The mean relative difference was 0.37 for AAC and 0.35 for ADC.

To ensure meaningful comparison, parameter settings were converted from T to T^* and S to S^* as explained in the Methods. Differences between AAC and ADC based on D-FHR and ta-FHR can be due to differences in the acquisition techniques, but may also be a result of unfair comparison due to intrinsic sampling errors. To investigate this, the ta-FHR traces were resampled to a 4 Hz signal using linear interpolation (data not shown). The equivalent analysis to the one in Figures 3 and 4, showed a correlation over 0.7 for T and S larger than 15 (0.25s). The correlation was decreasing with decreasing values for the parameters T and S (data not shown). These results are in agreement with the results reported in this paper using the original ta-FHR signals without interpolation.

In particular, the correlation between AAC (and ADC) based on D-FHR and the 4Hz resampled ta-FHR for the parameter settings presented in the literature are slightly different to the correlations found with the ta-FHR traces that were not resampled (Table 3). However the differences in correlation were very small and did not exceed an absolute value of 0.12 (mean = 0.04, standard deviation = 0.03).

Discussion

Main findings

Even though it has been suggested that PRSA analysis should preferably be performed on ta-FHR (Lobmaier et al. 2012), this is the first study investigating the actual impact of the acquisition technique on PRSA analysis. We applied PRSA analysis (average accelerative capacity, AAC, and average decelerative capacity, ADC) on simultaneous recordings using standard Doppler FHR (D-FHR) and transabdominal, ta-FHR (AN24, Monica Healthcare). We used 28 recording pairs acquired prior to labour and split them into two groups according to how similar the ta-FHR was with D-FHR (measured by the Root Mean Square error, RMS). We examined the correlation between PRSA on ta-FHR and PRSA on D-FHR across the full spectrum of possible AAC and ADC parameters values. As can be expected, the correlation between PRSA analysis on ta-FHR and D-FHR was higher in the subset of pairs with lower RMS error, i.e. the more similar the FHR between devices, the more similar the PRSA analysis. The correlation decreased with decreasing values of the PRSA parameters T and S: it was >0.8 for T and $S > 40$ (0.25s) gradually decreasing down to 0.1 for S below 5 (0.25s). Lower T values were also associated with decreased correlation but this was still above 0.5 when $S > 10$ (0.25s) (Figure 3 and Figure 4). We observed a small systematic difference (Fig. 7 and Fig. 8) which was not consistent across different values for S and T. For $T=5$ (0.25 s) and $S=45$ (0.25s), AAC computed from D-FHR have on average 0.3bpm higher values than the corresponding AAC computed from ta-FHR; similarly ADC computed from D-FHR have on average 0.3bpm lower values than the corresponding ADC computed from ta-FHR. For $T=S=3$ (0.25s), the systematic difference was larger (which is inline with the lower correlation in this case) and in the opposite direction and was -0.8 bpm and 0.8bpm for AAC and ADC respectively. We conclude that any systematic difference will need to be taken into account in future clinical studies and be considered specifically in the context of different algorithm parameter values.

PRSA analysis is a promising technique for clinical practice, but is still a research technique. We observed no differences between IUGR and controls in terms of their PRSA analysis neither on ta-FHR

nor on D-FHR. Our results demonstrate that future progress with the implementation of PRSA analysis needs to take into account the acquisition technique used.

Strengths and limitations

A strength of this study is the comprehensive spectrum of possible T and S values (between 1 and 50) that we examined as well as the detailed investigation of the parameters reported previously, as determined by our literature review. We studied both the AAC and ADC components of PRSA. A possible limitation of this study was a simplified formula for converting T and S to T* and S* (this assumed an average heart rate of 144 bpm). However, when considering the mesh plots (Figure 3 and Figure 4) one can see that the correlation mesh graph is fairly smooth, therefore it can be expected that the correlation will not vary significantly if T converted to slightly bigger or smaller T*.

The general trend of correlation across parameter values (Figures 3 and 4), was repeated in the resampled 4 Hz ta-FHR indicating that, despite differences in sampling frequency, there is a lack of correlation between high-resolution (low T and S values) PRSA analysis based on D-FHR and on ta-FHR. This further underlines the need to consider which acquisition technique is used in future analysis.

We did not exclude FHR decelerations or accelerations from analysis in-line with reports in the literature since 2013 (Table 2). As we analyzed only antenatal traces, decelerations were rare. But importantly, there is no universal method to exclude decelerations and accelerations. Published studies on PRSA that excluded them from analysis (Table 2) did not specify their methods for exclusion and therefore cannot be replicated. Moreover, methods to assign baseline and subsequently detect and exclude deviations, can vary significantly and thus provide considerably different outputs. As the purpose of this study was to examine the differences between PRSA on ta-FHR vs D-FHR we believe including the entire signal was appropriate.

A limitation of this study is that the gestational age of cases and controls was more than 34 weeks, while the gestational ages reported by some other authors was lower (Huhn & Lobmaier 2011; Stampalija et al. 2015; Graatsma et al. 2012; Fanelli et al. 2013; Fanelli et al. 2014). PRSA analysis on

both ta-FHR and D-FHR showed poor discriminatory power in our study. But it may be that CTG analysis beyond 34 weeks is a poor discriminator of IUGR and healthy babies (Gonçalves et al. 2013). In this context it should be noted that CTG analysis is primarily a dynamic test of fetal wellbeing and should be judged as such. Therefore, no reliable conclusions could be drawn from our study as to whether ta-FHR can provide more valuable information than D-FHR for the problem of IUGR discrimination (Baschat 2010). Additionally the sample size (28 registration pairs) was insufficient to draw strong conclusions regarding the discriminative capacity of PRSA analysis between IUGR and non-IUGR fetuses at these gestational ages (≥ 34 weeks).

All recordings in the study were taken prior to labour. Therefore the PRSA values were relatively low (between -4bpm and 4bpm). We speculate that our conclusions will be valid for recordings taken in labour where higher absolute PRSA values are more common, but further research is needed.

We compared the correlation between PRSA analysis of ta-FHR and D-FHR with the correlation between PRSA analysis of resampled ta-FHR (4Hz) and D-FHR. As we observed similar trends when comparing the correlations, this implies that the reported differences between PRSA analysis based on ta-FHR and D-FHR are most likely caused by a difference in acquisition techniques and not by sampling errors.

Conclusion

The acquisition technique does influence PRSA analysis especially for high resolution PRSA (low values of T or S) where there is low correlation between D-FHR PRSA and ta-FHR PRSA. Based on this dataset we were unable to differentiate between healthy and IUGR fetuses for both acquisition techniques (likely due to the larger gestational age of the fetuses).

The finding that the correlation between PRSA on ta-FHR and D-FHR is decreasing with lower values of T and S (i.e. high resolution PRSA) is interesting and important. It leads to two conclusions: (1) results from research that uses transabdominal monitors and PRSA parameters T and S below five (e.g. Rivolta et al., Stampalija et al) are not transferable to standard monitors; (2) the lack of correlation

specifically for small values of T and S indicate that the analysis of 'true beat to beat' data may provide different (and hopefully more useful) information when compared to standard FHR monitoring. This hypothesis remains to be tested in future work.

Acknowledgements

We would like to thank Gjerrit Meinsma of the Faculty of Electrical Eng., Mathematics and Computer Sciences of the Department of Applied Mathematics of the University of Twente for his supervision of J.A.M. van Scheepen.

Bibliography

- Abdulhay, E.W. et al., 2014. Review Article: Non-Invasive Fetal Heart Rate Monitoring Techniques. *Science, Biomedical*, 2(3), pp.53–67.
- Alfirevic, Z., Devane, D. & Gyte, G.M.L., 2013. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. *Cochrane database of systematic reviews (Online)*, 3(5), p.CD006066.
- Baschat, A.A., 2010. Fetal growth restriction – from observation to intervention. *Journal of Perinatal Medicine*, 38(3), pp.239–246. Available at: <http://www.degruyter.com/view/j/jpm.2010.38.issue-3/jpm.2010.041/jpm.2010.041.xml>.
- Bauer, A. et al., 2006. Phase-rectified signal averaging detects quasi-periodicities in non-stationary data. *Physica A: Statistical Mechanics and its Applications*, 364, pp.423–434. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S037843710501006X> [Accessed March 31, 2014].
- Bland, J.M. & Altman, D.G., 1986. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *The Lancet*, February 8, pp.307–310.
- Cesarelli, M. et al., 2007. An algorithm for the recovery of fetal heart rate series from CTG data. *Computers in Biology and Medicine*, 37(5), pp.663–669. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0010482506000886>.
- Cohen, W.R. et al., 2012. Accuracy and reliability of fetal heart rate monitoring using maternal abdominal surface electrodes. *Acta Obstetrica et Gynecologica Scandinavica*, 91(August), pp.1306–1313.
- Fanelli, A. et al., 2013. Quantitative Assessment of Fetal Well-Being Through CTG Recordings: A New Parameter Based on Phase-Rectified Signal Average. *IEEE Journal of Biomedical and Health Informatics*, 17(5), pp.959–966. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6530599> [Accessed March 5, 2014].

- Fanelli, A., Magenes, G. & Signorini, M.G., 2014. New Indices Extracted from Fetal Heart Rate Signal for the Assessment of Fetal Well-Being. In L. M. Roa Romero, ed. *IFMBE Proceedings: XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*. IFMBE Proceedings. Cham: Springer International Publishing, pp. 823–826. Available at: <http://link.springer.com/10.1007/978-3-319-00846-2> [Accessed February 19, 2014].
- Georgieva, A. et al., 2014. Phase-rectified signal averaging for intrapartum electronic fetal heart rate monitoring is related to acidaemia at birth. *BJOG: An International Journal of Obstetrics & Gynaecology*, p.n/a–n/a. Available at: <http://doi.wiley.com/10.1111/1471-0528.12568> [Accessed March 4, 2014].
- Gonçalves, H. et al., 2015. Comparison of the effect of different sampling modes on computer analysis of cardiotocograms. *Computers in Biology and Medicine*, 64, pp.62–66. Available at: <http://dx.doi.org/10.1016/j.compbiomed.2015.06.011>.
- Gonçalves, H. et al., 2006. Internal versus external intrapartum foetal heart rate monitoring : The effect on linear and nonlinear parameters. *physiological measurement*, (27).
- Gonçalves, H., Bernardes, J. & Ayres-de-campos, D., 2013. Gender-specific heart rate dynamics in severe intrauterine growth-restricted fetuses. *Early Human Development*, 89(6), pp.431–437. Available at: <http://dx.doi.org/10.1016/j.earlhumdev.2013.01.001>.
- Graatsma, E. et al., 2009. Fetal electrocardiography: feasibility of long-term fetal heart rate recordings. *BJOG: An International Journal of Obstetrics & Gynaecology*, 116(2), pp.334–338. Available at: <http://doi.wiley.com/10.1111/j.1471-0528.2008.01951.x>.
- Graatsma, E. et al., 2010. Maternal Body Mass Index Does not Affect Performance of Fetal Electrocardiography. *American Journal of Perinatology*, 27, pp.573–7.
- Graatsma, E.M. et al., 2012. Average acceleration and deceleration capacity of fetal heart rate in normal pregnancy and in pregnancies complicated by fetal growth restriction. *The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians*, 25(12), pp.2517–22. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22725720> [Accessed March 6, 2014].
- Hasan, M. a. et al., 2009. Detection and Processing Techniques of FECG Signal for Fetal Monitoring. *Biological Procedures Online*, 11(1), pp.263–295. Available at: <http://www.biologicalproceduresonline.com/content/11/1/263>.
- Huhn, E. & Lobmaier, S., 2011. New computerized fetal heart rate analysis for surveillance of intrauterine growth restriction. *Prenatal Diagnosis*, (December 2010), pp.509–514. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/pd.2728/full> [Accessed March 5, 2014].
- Lobmaier, S.M. et al., 2012. Phase-rectified signal averaging as a new method for surveillance of growth restricted fetuses. *The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians*, 25(12), pp.2523–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22630786> [Accessed March 5, 2014].

- Piéri, J.F. et al., 2001. Compact long-term recorder for the transabdominal foetal and maternal electrocardiogram. *Medical & Biological Engineering & Computing*, 39(1), pp.118–125. Available at: <http://link.springer.com/10.1007/BF02345275>.
- Reinhard, J. et al., 2012. Intrapartum signal quality with external fetal heart rate monitoring: A two way trial of external Doppler CTG ultrasound and the abdominal fetal electrocardiogram. *Archives of Gynecology and Obstetrics*, 286(5), pp.1103–1107.
- Rivolta, M.W. et al., 2014. Acceleration and Deceleration Capacity of Fetal Heart Rate in an In-Vivo Sheep Model D. Hodgson-Zingman, ed. *PLoS ONE*, 9(8), p.e104193. Available at: <http://dx.plos.org/10.1371/journal.pone.0104193>.
- Sänger, N. et al., 2012. Prenatal Foetal Non-invasive ECG instead of Doppler CTG - A Better Alternative? *Geburtshilfe und Frauenheilkunde*, 72(07), pp.630–633. Available at: <http://www.thieme-connect.de/DOI/DOI?10.1055/s-0032-1315012>.
- Signorini, M.G., Fanelli, A. & Magenes, G., 2014. Monitoring Fetal Heart Rate during Pregnancy: Contributions from Advanced Signal Processing and Wearable Technology. *Computational and Mathematical Methods in Medicine*, 2014, pp.1–10. Available at: <http://www.hindawi.com/journals/cmmm/2014/707581/> [Accessed March 5, 2014].
- Stampalija, T. et al., 2015. Parameters influence on acceleration and deceleration capacity based on trans-abdominal ECG in early fetal growth restriction at different gestational age epochs. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 188, pp.104–112. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0301211515000743>.
- Tagliaferri, S. et al., 2015. Evaluation of the Acceleration and Deceleration Phase-Rectified Slope to Detect and Improve IUGR Clinical Management. *Computational and Mathematical Methods in Medicine*, 2015.