Received Date : 29-Jul-2016
Revised Date : 14-Nov-2016
Accepted Date : 16-Nov-2016
Article type : Original Research Article

Agreement and accuracy using the FIGO, ACOG and NICE cardiotocography interpretation guidelines

Running headline Comparison of three CTG guidelines

Susana SANTO[1], Diogo AYRES-DE-CAMPOS[2], Cristina COSTA-SANTOS[3], William SCHNETTLER[4], Austin UGWUMADU[5], Luís MENDES DA GRAÇA[1], for the FM-COMPARE collaboration*

[1]Department of Obstetrics and Gynecology , Santa Maria Hospital, Faculty of Medicine of Lisbon University, Lisbon,
[2]Department of Obstetrics and Gynecology, Medical School, University of Porto, S. João Hospital, Institute of Biomedical Engineering, Porto**,**
[3]Department of Medical Informatics, Medical School, University of Porto, Porto, Portugal
[4]Center for Maternal Cardiac Care, TriHealth, Good Samaritan Hospital, Cincinnati, OH, USA
[5]Department of Obstetrics & Gynaecology, St George's Hospital, University of London, London, UK

*Members of the FM-COMPARE collaboration: Nuno Clode, Luisa Pinto, Monica Centeno, Rui Carvalho, Pedro Rocha, Inês Reis, Maria Afonso, Catarina Castro, Ralph Aserkoff, Toni Golen, Hope Ricciotti, Mary Vadnais, Jennifer Scott, Sarosh Rana, Hayley Quant, Lynsey Caldwell, Edwin Chandraharan, Amar Bhide, Nicola Lack, Latika Narang, Sadia Muhammad, Sangeetha Devarajan, Rachel Squires, Karolina Affors, Lilian Ugwumadu.

Corresponding Author:
Susana Santo
Departamento de Obstetrícia e Ginecologia, Hospital Santa Maria, Avenida Professor Egas Moniz, 1649-035 Lisboa, Portugal
Email: sfsanto@gmail.com

Conflicts of Interest: The authors report no conflicts of interest.

Abstract

Introduction: One of the limitations reported with cardiotocography (CTG) is the modest interobserver agreement observed in tracing interpretation. This study compared agreement, reliability and accuracy of CTG interpretation using the FIGO, ACOG and NICE guidelines. *Material and methods:* A total of 151 tracings was evaluated by 27 clinicians from three centers where FIGO, ACOG and NICE guidelines were routinely used. Interobserver agreement was evaluated using the proportions of agreement (PA) and reliability with the kappa (k) statistic. The accuracy of tracings classified as "pathological/category III" was assessed for prediction of newborn acidemia. For all measures, 95% confidence intervals (95%CI) were calculated. *Results:* CTG classifications were more distributed with FIGO (9%, 52%, 39%) and NICE (30%, 33%, 37%) than with ACOG (13%, 81%, 6%). The category with the highest agreement was ACOG category II (PA=0.73 95%CI 0.70-76), and the ones with the lowest agreement were ACOG categories I and III. Reliability was significantly higher with FIGO (k=0.37, 95%CI 0.31-0.43), and NICE (k=0.33, 95%CI 0.28-0.39) than with ACOG (k= 0.15, 95%CI 0.10-0.21), however all represent only slight/fair reliability. FIGO and NICE showed a trend towards higher sensitivities in prediction of newborn acidemia (89% and 97% respectively) than ACOG (32%,), but the latter achieved a significantly higher specificity (95%). *Conclusions:* With ACOG guidelines there is high agreement in category II, low reliability, low sensitivity and high specificity in prediction of acidemia. With FIGO and NICE guidelines there is higher reliability, a trend towards higher sensitivity, and lower specificity in prediction of acidemia.

Keywords
cardiotocography, heart rate, electronic fetal monitoring, guidelines, agreement

Key message

Agreement, reliability and accuracy of cardiotocography interpretation using the FIGO, ACOG and NICE guidelines are compared. The study demonstrates significant differences between these three major classification systems that are important for the development of future guidelines.

Abbreviations

ACOG - American College of Obstetrics and Gynecology
CTG - cardiotocography
FHR - fetal heart rate
FIGO - International Federation of Gynecology and Obstetrics
NICE - National Institute of Health and Clinical Excellence
PA - proportions of agreement

Introduction

Cardiotocography (CTG) is an integral part of intrapartum care in most high-income countries. However, one of its limitations is the modest interobserver agreement in CTG interpretation (1–5). The International Federation of Gynecology and Obstetrics (FIGO) published its first guidelines on fetal heart rate (FHR) monitoring in 1987 (6) and established the only international consensus available at the time the present study was undertaken. Many national scientific organizations have also published guidelines on the subject, but perhaps those with the largest impact were developed by the American College of Obstetricians and Gynecologists (ACOG) and the United Kingdom National Institute of Clinical Excellence (NICE). ACOG has published several revised versions of their original publication in 1974 (7), the last of which in 2010 in association with the National Institute of Child Health and Human Development and the Society for Maternal-Fetal Medicine (8). The Royal College of Obstetricians and Gynaecologists published its first guidelines in 2001, and updated them in 2007 in association with NICE (9). This was the latest version available at the time the present study was undertaken.

These three guidelines have important differences, not only in the definition of individual CTG features but also in the criteria used for overall tracing classification (tables 1 and 2)(10). The aim of this study was to compare interobserver agreement, reliability and accuracy of CTG analysis, when performed according to the FIGO, ACOG and NICE guidelines. The hypothesis was that the differences in guideline structure, as well as in clarity and complexity of definitions, could result in different interobserver agreements, and in different predictive capacities for CTG interpretation. A second hypothesis was that observer experience would have an additional impact on these findings.

Material and methods

Cases were selected from a pre-existing database of intrapartum CTGs acquired in a tertiary-care university hospital (11). All patients gave their written informed consent for their tracings to be used in an anonymous way for research purposes. Laboring women were consecutively selected if they fulfilled the following inclusion criteria: singleton pregnancy, ≥ 37 gestational weeks, fetus in cephalic presentation, absence of known fetal malformations, active phase of labor and an established indication for continuous CTG monitoring (augmented or induced labor, meconium staining of the amniotic fluid, abnormalities detected on admission CTG or on intermittent fetal auscultation). All patients were continuously monitored until delivery, using a fetal electrode and an external tocodynamometer.

Paired umbilical cord blood sampling and analysis was performed in all cases, and fetal acidemia was defined as an umbilical artery pH value of 7.05 or less. Cases were subsequently excluded if one of the following situations was documented: total tracing length less than 60 minutes, signal loss in the last hour of the tracing exceeding 15%, interval between tracing-end and vaginal birth exceeding 5 minutes, or interval between tracing-end and cesarean birth exceeding 20 minutes, complications with the potential to influence fetal oxygenation recorded between tracing-end and delivery (shoulder dystocia, difficult cesarean extraction, etc), anesthetic complications at the time of delivery, or invalid cord blood gas values (11).

A total of 193 patients were enrolled and 42 were subsequently excluded, leaving 151 cases for analysis in the study. Only the last 60 minutes of patients' tracings obtained before delivery were presented to clinicians. No additional clinical information was provided, except

that records were acquired just before birth in singleton term pregnancies. CTG tracings were presented at a paper speed of 1 cm/ min to the group of clinicians using the FIGO and NICE guidelines and at a paper speed of 3 cm/min to the group of clinicians using the ACOG guidelines.

A total of 27 clinicians performed the analysis of CTGs, nine from each of three different centers where the referred guidelines were routinely used. The FIGO guidelines group were recruited from the Santa Maria Hospital in Lisbon, Portugal, the ACOG guidelines group from the Beth Israel Deaconess Medical Centre in Boston, USA, and the NICE guidelines group from St. George's Hospital - University of London, UK. At each center, three of the selected clinicians had more than 10 years of experience in CTG analysis, three had six to10 years of experience, and three had less than six years of experience. Each clinician only evaluated the 151 tracings once. and according to the guidelines he/she was accustomed to.

Clinicians received digital copies of the tracings by email in Word format, together with a file summarizing the main points of the guidelines to be used. They were asked to view the tracings independently and to evaluate FHR baseline, variability, accelerations and decelerations, before attributing an overall tracing classification.

Statistical analyses

Interobserver agreement was assessed using the proportions of agreement (PA) and the proportion of specific agreement (PA for each category), as recommended by the "Guidelines for reporting reliability and agreement studies" (GRRAS) (12). For all results, 95% confidence intervals were calculated, and findings were considered significantly different if these intervals did not overlap. If the lower limit of the 95% confidence intervals for PA was under 0.50, agreement was also considered to be poor (13). Reliability was evaluated with the kappa statistic (k-Light's kappa for n raters), which adjusts PA to the agreement expected by chance, so the distribution of ratings in the different classes influences the results. It is possible to obtain a high PA and a low kappa when the prevalence of a given rating is very high or low (14). Kappa values below 0.20 were considered as slight reliability; those ranging between 0.21 and 0.40 as fair reliability, those between 0.41 and 0.60 as moderate reliability, those between 0.61 and 0.80 as substantial reliability, and values larger than 0.80 as almost perfect reliability (15). Tracings classified as pathological/category

III were compared with all the others regarding their capacity to predict newborn acidemia. Sensitivity and specificity were calculated with 95% of confidence intervals. Statistical analysis was performed using the R package obs.agree version 1.0 (Free Software Foundation, Boston, USA).

Approval by the S. João Hospital, Porto Medical School Research Ethics Committee was obtained for the study (Comissão de Ėtica do Centro Hospitalar de São João, Parecer N ° 28/2010, 19/07/2010). Procedures were in accordance with the ethical standards of the committee and with the Helsinki Declaration.

Results

All tracings were analyzed by the 27 clinicians, for a total of 4077 evaluations. Table 3 displays the evaluation of basic CTG features and overall tracing classification by clinicians in each study group. In all groups, the majority of tracings were evaluated as having normal baseline and normal variability. Clinicians in the FIGO and ACOG groups considered that most tracings had accelerations, while those in the NICE group considered the opposite. All groups identified decelerations in the majority of tracings. The ACOG group classified 81% of tracings as category II, while the suspicious classification was only selected by 52% in the FIGO group and 33% in the NICE group.

Interobserver agreement and reliability in evaluation of basic CTG features and overall tracing classification are displayed in table 4. For FHR baseline, agreement and reliability were high and similar in all groups. The highest agreement was achieved in identification of a normal FHR baseline, and results were significantly better in the ACOG and NICE groups than in the FIGO group. Bradycardia showed the lowest agreement, and no differences between the groups were identified.

A high agreement was found in the evaluation of variability, with no significant differences occurring between the groups. All groups showed the highest agreement in identification of normal variability.

For identification of accelerations, a similar agreement was found between all groups, with the NICE group showing a significantly higher agreement in identification of “no accelerations”.

The FIGO group had a higher agreement than ACOG in identification of decelerations (both present and absent), and all groups showed a poorer agreement in identification of absent decelerations.

In overall tracing classification, the ACOG group had a significantly higher agreement than FIGO, and both had a significantly higher agreement than NICE. While in the ACOG group category II classification reached a significantly higher agreement than any other guideline classification, category I and category III obtained a significantly lower agreement than others. A significantly lower reliability was obtained with the ACOG classification than with FIGO or NICE. Kappa values in overall tracing classification, represent a slight/fair reliability with all guidelines.

Table 5 displays interobserver agreement according to the number of years of experience in CTG analysis. Clinicians with less than six years of experience in the ACOG group showed the highest agreement in tracing classification, but this was mainly due to agreement on category II. In the FIGO and NICE groups there were no significant differences in agreement, according to the level of experience.

In the 151 cases evaluated there were seven newborns with an umbilical artery blood pH $\leq$ 7.05, but no cases of hypoxic-ischemic encephalopathy. The sensitivity and specificity of category III/ pathological tracings in prediction of acidemia is displayed in table 6. The FIGO and NICE groups showed a trend towards a higher sensitivity than ACOG, but the differences were not statistically significant. On the other hand, the ACOG group showed a significantly higher specificity than the others. No significant differences were found in these comparisons between levels of expertise.

Discussion

This study compares the agreement, reliability and accuracy of FIGO, ACOG and NICE guidelines for CTG interpretation and showed that attribution of category II is very frequent with the ACOG guidelines, leading to a high overall interobserver agreement, a low reliability, a low sensitivity and a high specificity of category III tracings in prediction of fetal acidemia. With the FIGO and NICE guidelines, a more balanced distribution of classifications is seen, and there appears to be a higher sensitivity and a lower specificity of pathological tracings in prediction of fetal acidemia.

This study also confirms that there is high agreement in identification of normal baseline, tachycardia, normal variability, presence of accelerations and decelerations. It was not possible to evaluate the classification of decelerations, as these events are defined differently in the three guidelines.

Other studies evaluating the reproducibility of CTG analysis using the FIGO and ACOG guidelines have shown that there is a fair to good agreement in evaluation of the baseline and accelerations, and a poor agreement regarding decelerations (3,16). It has also been reported that CTG classification as category I/normal is more reproducible than the other categories (4,5,17–19). Our study demonstrates that this depends on the selected guidelines. With the ACOG guidelines, classification in category I it was less reproducible, while with the other guidelines differences were usually small and not statistically significant.

Three-tiered classification systems usually suggest no action for category I/normal tracings and rapid intervention for category III/pathological tracings. Thus, these two categories are probably the ones more directly associated with outcomes and interventions rates. A low percentage of tracings considered normal may be associated with higher rate of obstetric intervention, while a low percentage of tracings considered pathological may be associated with poor neonatal outcomes. Category II/suspicious includes a broad spectrum of heterogeneous FHR patterns that are inconsistently associated with fetal acidemia, making clinical management of these situations more uncertain.

Several studies have shown that CTG has a high sensitivity and a limited specificity in the prediction of fetal hypoxia/acidosis. Our study demonstrates that this finding depends on the interpretation guidelines used. The ACOG guidelines tended to classify abnormal patterns more in category II, because of more restrictive criteria for category III, and some acidemia cases were classified in category II, hence the tendency for a lower sensitivity and higher specificity of these guidelines. With the FIGO and NICE guidelines acidemia cases were more in the pathological category, thereby increasing sensitivity for the detection of acidemia but decreasing specificity. These results, however, need to be interpreted with caution, given the low number of cases with newborn acidemia.

Interobserver agreement and accuracy were not majorly affected by clinicians' years of practice for FIGO and NICE groups, suggesting that they can be generalized to all clinicians with at least 6 years of experience. Similar findings have also been reported by others (16,19–22). On the other hand, clinicians with less years of experience may follow the guidelines more strictly and this may be responsible for the slightly better agreement obtained in the ACOG group.

The main strengths of the study are that it involved a large number of clinicians working in different centers where the CTG guidelines were routinely used, with paper speeds they were accustomed to. The selection of different years of clinical experience also contributes to a greater generalizability of results. In selection of tracings, only cases monitored until very close to birth were included, so that umbilical artery pH would closely reflect fetal hypoxia/acidosis occurring during the last minutes of labor.

The number of cases selected for analysis was decided somewhat empirically, taking into account the expected capacity of observers to complete the task within a reasonable time period, and given the modest number of cases with acidemia in this sequentially selected population, it resulted in large confidence intervals for the sensitivity analysis. Tracing analysis was carried out at leisure, with immediate access to the guidelines, and the full 60-minute tracings were made available. These conditions are very different from daily practice, where time pressure, memory recall of the guidelines, and frequent re-evaluation of ongoing tracings are the norm. The immediate availability of guidelines removes the memory issues that may be involved in tracing interpretation and focuses more on clinicians' capacity to identify patterns and to follow guidance.

Centers were selected because they used the referred guidelines in routine clinical practice, but the possibility of local or even individual adaptation of the guidelines cannot be ruled out, as well as the effect of local training and audit. All centers carry out regular CTG training, but course frequencies and methodologies are different. Local culture may for instance have been responsible for the decreased number of accelerations identified in the NICE guideline group, as the more rounded increases in FHR occurring after decelerations were most likely considered "shoulders" and not "true" accelerations. The last 60 minutes of the tracing were evaluated, as similar periods are commonly used for tracing classification, but the initial part may have been different from the end, and clinicians may have evaluated

this in different ways. The period before birth is usually the most challenging for CTG interpretation, and agreement could have been different in a more stable period of labor. To ensure a reasonable signal quality, internal FHR monitoring was used in all cases, but again a different agreement could have been achieved with external monitoring and greater signal loss. The sequential selection of cases with subsequent exclusion criteria guarantees the generalizability of results to a population that has good signal quality tracings and no unmonitored hypoxic events, but this does not occur in all intrapartum cases. It also resulted in a low number of cases with newborn acidemia, with consequences on the robustness of the accuracy analysis.

This study shows that there are important differences in the way clinicians interpret CTG tracings, depending on the guidelines they use. Differences in guideline structure, as well as in clarity and complexity of definitions, have a profound effect on interobserver agreement and reliability, as well as on the sensitivity and specificity of CTG classifications in predicting acidemia. These aspects need to be taken into consideration in the development of new guidelines.

References

1. Borgatta L, Shrout PE, Divon MY. Reliability and reproducibility of nonstress test readings. Am J Obstet Gynecol. 1988;159:554–558.

2. Donker DK, Van Geijnb HP, Hasman A. Interobserver variation in the assessment recordings of fetal heart rate. Eur J Obstet Gynecol Reprod Biol. 1993;52:21–28.

3. 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;57:33–37.

4. Ayres-de-Campos D, Bernardes J, Costa-Pereira A, Pereira-Leite L. Inconsistencies in classification by experts of cardiotocograms and subsequent clinical decision. Br J Obstet Gynaecol. 1999;106:1307–1310.

5. Blackwell SC, Grobman WA, Antoniewicz L, Hutchinson M, Gyamfi Bannerman C. Interobserver and intraobserver reliability of the NICHD 3-Tier Fetal Heart Rate Interpretation System. Am J Obstet Gynecol. 2011;205:378.e1–5.

6. FIGO subcommittee on Standards on Perinatal Medicine. Guidelines for the use of fetal monitoring. Int J Gynaecol Obstet. 1987;25:159–167.

7. American College of Obstetricians and Gynecologists. Fetal heart rate monitoring: guidelines. ACOG Tech Bull 32. 1974;June.

8. American College of Obstetricians and Gynecologists. Management of intrapartum fetal heart rate tracings. Practice Bulletin 116: Obstet Gynecol. 2010;116:1232–1240.

9. National Collaborating Centre for Women's and Children's Health commissioned by the National Institute for Health and Clinical Excellence. Intrapartum care. 2007.

10. Ayres-de-Campos D, Bernardes J. Twenty-five years after the FIGO guidelines for the use of fetal monitoring: time for a simplified approach? Int J Gynaecol Obstet. 2010;110:1–6.

11. Costa A, Ayres-de-Campos D, Costa F, Santos C, Bernardes J. Prediction of neonatal acidemia by computer analysis of fetal heart rate and ST event signals. Am J Obstet Gynecol. 2009;201:464.e1–6.

12. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64:96–106.

13. Grant JM. The fetal heart rate trace is normal, isn't it? Observer agreement of categorical assessments. Lancet. 1991;337:215–218.

14. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43:543–549.

15. Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–174.

16. Devoe L, Golde S, Kilman Y, Morton D, Shea K, Waller J. A comparison of visual analyses of intrapartum fetal heart rate tracings according to the new national institute of child health and human development guidelines with computer analyses by an automated fetal heart rate monitoring system. Am J Obstet Gynecol. 2000;183:361–366.

17. Beaulieu MD, Fabia J, Leduc B, Brisson J, Bastide A, Blouin D et al. The reproducibility of intrapartum cardiotocogram assessments. Can Med Assoc J. 1982;127:214–216.

18. Blix E, Sviggum O, Koss KS, Øian P. Inter-observer variation in assessment of 845 labour admission tests: Comparison between midwives and obstetricians in the clinical setting and two experts. BJOG. 2003;110:1–5.

19. Schiermeier S, Westhof G, Leven A, Hatzmann H, Reinhard J. Intra- and Interobserver Variability of Intrapartum Cardiotocography : A Multicenter Study Comparing the FIGO Classification with Computer Analysis Software. Gynecol Obstet Invest. 2011;72:169–73.

20. Lidegaard O, Bøttcher LM , Weber T. Description, evaluation and clinical decision making according to various fetal heart rate patterns. Inter-observer and regional variability. Acta Obstet Gynecol Scand. 1992;71:48–53.

21. Palomäki O, Luukkaala T, Luoto R, Tuimala R. Intrapartum cardiotocography – the dilemma of interpretational variation. J Perinat Med. 2006;34:298–302.

22. Epstein AJ, Twogood S, Lee RH, Opper N, Beavis A, Miller DA. Interobserver reliability of fetal heart rate pattern interpretation using NICHD definitions. Am J Perinatol. 2013;30:463–468.

Table 1. Comparison of the basic fetal heart rate (FHR) definitions contained in the International Federation of Gynecology and Obstetrics (FIGO) guidelines of 1987, the American College of Obstetricians and Gynecologists (ACOG) guidelines of 2010 and the National Institute of Clinical Excellence (NICE) guidelines of 2007.

BASELINE	FIGO	Baseline FHR is the mean level of the FHR when this is stable, accelerations and decelerations being absent. It is determined over a time period of 5 or 10 min and expressed in bpm.
	NICE	Mean level of the FHR when this stable, excluding accelerations and decelerations. It is determined over a time period of 5 or 10 minutes and expressed in bpm
	ACOG	Mean FHR rounded to increments of 5 bpm during a 10-minute segment, excluding: periodic or episodic changes, periods of marked FHR variability, segments of baseline that differ > 25 bpm
Normal baseline	FIGO	110-150 bpm
	NICE	110-160 bpm
	ACOG	110-160 bpm
Tachycardia	FIGO	(no definition)
	NICE	>180 bpm (161-180 bpm is moderate tachycardia)
	ACOG	>160 bpm
Bradycardia	FIGO	< 80 bpm
	NICE	<100 bpm (100-109 bpm is moderate bradycardia)
	ACOG	<110 bpm
VARIABILITY	FIGO	Oscillations of FHR around its mean level (long-term variability). This is usually only quantitated by description of the amplitude of the oscillations around the baseline heart rate.
	NICE	The minor fluctuations in baseline FHR occurring at three to five cycles per minute. It is measured by estimating the difference in bpm between the highest peak and lowest trough of fluctuation in a one-minute segment of the trace
	ACOG	Fluctuations in the baseline FHR that are irregular in amplitude and frequency. It is visually quantitated as the amplitude of peak-to-through in bpm.
Normal variability	FIGO	Between 5-25 bpm
	NICE	Greater than or equal to 5 bpm between contractions
	ACOG	Amplitude range 6-25 bpm (moderate variability)
Reduced variability	FIGO	< 5 bpm for more than 40 minutes (suspicious if variability 5-10 bpm for more than 40 minutes)
	NICE	Less than 5 bpm for 40-90 minutes (non-reassuring) or >90 minutes (abnormal variability)
	ACOG	Amplitude range 5 bpm or fewer (minimal variability)
Increased variability	FIGO	> 25 bpm
	NICE	-
	ACOG	Amplitude range greater than 25 bpm (marked variability)
ACCELERATIONS	FIGO	Transient increase in heart rate of 15 bpm or more and lasting 15 seconds or more.
	NICE	Transient increases in FHR of 15 bpm or more and lasting 15 seconds or more.
	ACOG	A visually apparent abrupt increase (onset to peak in less than 30 seconds) in the FHR. Beyond 32 weeks of gestation, an acceleration has a peak of 15 bpm or more above the baseline, with a duration of 15 seconds or more but less than 2 minutes from onset to return. Prolonged accelerations last 2 minutes or more but less than 10 minutes.
DECELERATIONS	FIGO	Transient episodes of slowing of FHR below the baseline level of more than 15 bpm and lasting 10 seconds or more.
	NICE	Transient episodes of slowing of FHR below the baseline level of more than 15 bpm and lasting 15 seconds or more.
	ACOG	-
Early decelerations	FIGO	-
	NICE	Uniform, repetitive, periodic slowing of FHR with onset early in the contraction and return to baseline at the end of the contraction.
	ACOG	Visually apparent usually symmetrical gradual decrease and return of the FHR associated with a uterine contraction. A gradual decrease is defined as from the onset to the FHR nadir of 30 seconds or more. The decrease in FHR is calculated from the onset to the nadir of the deceleration. The nadir of the deceleration occurs at the same time as the peak of the contraction. In most cases, the onset, nadir, and recovery of the deceleration are coincident with the beginning, peak, and ending of the contraction, respectively.
Late decelerations	FIGO	-
	NICE	Uniform, repetitive, periodic slowing of FHR with onset mid to end of the contraction and nadir more than 20 seconds after the peak of the contraction and ending after the contraction. In the presence of a non-accelerative trace with baseline variability less than 5 bpm, the definition would include decelerations less than 15 bpm
	ACOG	Visually apparent usually symmetrical gradual decrease and return of the FHR associated with a uterine contraction. A gradual decrease is

		defined as from the onset to the FHR nadir of 30 seconds or more. The decrease in FHR is calculated from the onset to the nadir of the deceleration. The deceleration is delayed in timing, with the nadir of the deceleration occurring after the peak of the contraction. In most cases, the onset, nadir, and recovery of the deceleration occur after the beginning, peak, and ending of the contraction, respectively.
Variable decelerations	FIGO	-
	NICE	Variable, intermittent periodic slowing of FHR with rapid onset and recovery. Time relationships with contraction cycle are variable and they may occur in isolation. Sometimes they resemble other types of deceleration patterns in timing and shape*.
	ACOG	Visually apparent abrupt decrease in FHR. An abrupt decrease is defined as from the onset of the deceleration to the beginning of the FHR nadir of less than 30 seconds. The decrease in FHR is calculated from the onset to the nadir of the deceleration. The decrease in FHR is 15 bpm or greater, lasting 15 seconds or greater, and less than 2 minutes in duration. When variable decelerations are associated with uterine contractions, their onset, depth, and duration commonly vary with successive uterine contractions.
Prolonged decelerations	FIGO	-
	NICE	An abrupt decrease in FHR to levels below the baseline that lasts at least 60-90 seconds. These decelerations become pathological if they cross two contractions (i.e. greater than 3 minutes)
	ACOG	Visually apparent decrease in the FHR below the baseline. Decrease in FHR from the baseline that is 15 bpm or more, lasting 2 minutes or more but less than 10 minutes in duration. If a deceleration lasts 10 minutes or longer, it is a baseline change.
SINUSOIDAL PATTERN	FIGO	Regular cyclic changes in the FHR baseline, such as the sine wave. The characteristics of the pattern being: the frequency is less than 6 cycles/min, the amplitude is at least 10 bpm and the duration should be 20 minutes or longer.
	NICE	A regular oscillation of the baseline long-term variability resembling a sine wave. This smooth, undulating pattern, lasting at least 10 minutes, has a relatively fixed period of 3-5 cycles per minute and an amplitude of 5-15 bpm above and below the baseline. Baseline variability is absent.
	ACOG	Visually apparent, smooth, sine wave-like undulating pattern in FHR baseline with a cycle frequency of 3-5 per minute which persists for 20 minutes or more.

*The NICE guidelines also define “atypical variable decelerations” when the following additional components are found: loss of primary or secondary rise in baseline rate, slow return to baseline FHR after the end of a contraction, prolonged secondary rise in baseline rate, byphasic deceleration, loss of variability during deceleration, continuation of baseline rate at lower level. Bpm, beats per minute. Adapted from Ayres-de-Campos D, Bernardes J. Twenty-five years after the FIGO guidelines for the use of fetal monitoring: time for a simplified approach? Int J Gynaecol Obstet. 2010;110(1):1–6.

Table 2. Comparison of cardiotocography (CTG) classification criteria in the FIGO, NICE and ACOG guidelines.

FIGO	NICE	ACOG
NORMAL PATTERN - Baseline heart rate between 110 and 150 bpm - Amplitude of heart rate variability between 5 and 25 bpm	**NORMAL** *(a CTG where all of the following four reassuring features are present)* - Baseline rate: 110-160 bpm - Variability: ≥5 bpm - No decelerations - Accelerations: present	**CATEGORY I** *(category I FHR tracings include all of the following)* - Baseline rate: 110-160 bpm - Baseline variability: 6-25 bpm - Late or variable decelerations: absent - Early decelerations: present or absent - Accelerations: present or absent
SUSPICIOUS PATTERN - Baseline heart rate between 150 and 170 bpm or between 100 and 110 bpm - Amplitude of variability between 5 and 10 bpm for more than 40 minutes - Increased variability above 25 bpm - Variable decelerations	**SUSPICIOUS** *(a CTG where **one** of the following features is present and all others fall into the reassuring category)* - Baseline rate 100-109 bpm 161-180 bpm - Baseline variability < 5 bpm for 40-90 minutes - Decelerations Typical variable decelerations with over 50% of contractions occurring for over 90 minutes Single prolonged deceleration for up to 3 minutes - Accelerations The absence of accelerations with an otherwise normal trace is of uncertain significance	**CATEGORY II** *(Category II FHR tracings include all FHR tracings not categorised as Category I or Category III. Examples of Category II FHR tracings include any of the following)* - Baseline rate Bradycardia not accompanied by absent baseline variability Tachycardia - Baseline variability Minimal variability Absent variability with no recurrent decelerations Marked variability - Accelerations Absence of induced accelerations after fetal stimulation - Periodic or episodic decelerations Recurrent variable decelerations accompanied by minimal or moderate baseline variability Prolonged deceleration 2-10 minutes Recurrent late decelerations with moderate baseline variability Variable decelerations with other characteristics such as slow return to baseline, overshoots or shoulders
PATHOLOGICAL PATTERN - Baseline heart rate below 100 or above 170 bpm - Persistence of heart rate variability of less than 5 bpm for more than 40 minutes - Severe variable decelerations or severe repetitive early decelerations. - Prolonged decelerations - Late decelerations: the most ominous trace is a steady baseline without baseline	**PATHOLOGICAL** *(a CTG with **one or more** of the following features **or two or more** features in the previous category)* - Baseline rate < 100 bpm > 180 bpm Sinusoidal pattern ≥ 10 minutes - Baseline variability < 5 bpm for ≥ 90 minutes	**CATEGORY III** *(Category III FHR tracings include either)* - Absent baseline FHR variability and any of the following: Recurrent late decelerations Recurrent variable decelerations Bradycardia - Sinusoidal pattern

variability and with small decelerations after each contraction - A sinusoidal pattern	- Decelerations atypical variable decelerations with over 50% contractions for > 30 minutes Late decelerations for > 30 minutes Prolonged deceleration > 3 minutes

FIGO, International Federation of Gynecology and Obstetrics; ACOG, the American College of Obstetricians and Gynecologists, NICE, National Institute of Clinical Excellence; FHR, fetal heart rate; bpm, beats per minute.
Adapted from Ayres-de-Campos D, Bernardes J. Twenty-five years after the FIGO guidelines for the use of fetal monitoring: time for a simplified approach? Int J Gynaecol Obstet. 2010;110(1):1–6.

Table 3. Distribution of the different evaluations of basic cardiotocography (CTG) features and overall tracing classification by the three groups of clinicians, using the FIGO, ACOG and NICE guidelines respectively, n = number of ratings, % = percentage of tracings where these ratings were attributed. Total number of CTG evaluations for each study group was 1359 (151 x 9 clinicians).

CTG features		**FIGO 1987**		**ACOG 2010**		**NICE 2007**	
		n	%	n	%	n	%
Normal	Normal	848	62	1120	82	1079	79
	Tachycardia	443	33	195	14	251	19
	Bradycardia	68	5	44	3	29	2
Variability	Normal	1066	78	1127	83	1157	85
	Abnormal	293	22	232	17	202	15
Accelerations	Present	816	60	805	59	529	39
	Absent	543	40	554	41	830	61
Decelerations –	Present	1228	90	1205	89	1207	89
	Absent	131	10	154	11	152	11
Overall tracing classification							
	Cat. I /Normal	116	9	171	13	401	30
	Cat. II /Suspicious	712	52	1106	81	452	33
	Cat. III /Pathological	531	39	82	6	506	37

FIGO, International Federation of Gynecology and Obstetrics; ACOG, the American College of Obstetricians and Gynecologists, NICE, National Institute of Clinical Excellence.

Table 4. Interobserver agreement evaluated by the proportions of agreement (PA), and reliability evaluated by the kappa statistics (k) with respective 95% confidence intervals (95% CI), for the evaluation of basic cardiotocography (CTG) features and overall tracing classification by the three study groups of clinicians.

	FIGO 1987		**ACOG 2010**		**NICE 2007**	
	PA (95% CI)	k (95% CI)	PA (95% CI)	k (95% CI)	PA (95% CI)	k (95% CI)
FHR baseline	**0.81 (0.78-0.85)**	**0.63 (0.57-0.70)**	**0.88 (0.84-0.91)**	**0.59 (0.49-0.69)**	**0.88 (0.85-0.91)**	**0.65 (0.58-0.72)**
Normal	0.86 (0.83-0.89)		0.93 (0.90-0.95)		0.93 (0.91-0.96)	
Tachycardia	0.80 (0.73-0.85)		0.67 (0.57-0.77)		0.73 (0.66-0.80)	
Bradycardia	0.40 (0.25-0.54)		0.49 (0.07-0.71)		0.42 (0.00-1.00)	
Variability	**0.83 (0.80-0.86)**	**0.51 (0.42-0.61)**	**0.85 (0.82-0.88)**	**0.49 (0.39-0.59)**	**0.83 (0.80-0.86)**	**0.38 (0.29-0.50)**
Normal	0.89 (0.87-0.92)		0.91 (0.89-0.93)		0.90 (0.88-0.92)	
Abnormal	0.61 (0.51-0.69)		0.57 (0.45-0.66)		0.44 (0.32-0.53)	
Accelerations	**0.67 (0.64-0.71)**	**0.34 (0.29-0.41)**	**0.67 (0.63-0.70)**	**0.34 (0.28-0.40)**	**0.71 (0.68-0.75)**	**0.41 (0.35-0.48)**
Yes	0.73 (0.68-0.77)		0.72 (0.67-0.76)		0.61 (0.57-0.68)	
No	0.59 (0.54-0.64)		0.59 (0.53-0.64)		0.76 (0.72-0.80)	
Decelerations	**0.92 (0.89-0.95)**	**0.53 (0.43-0.66)**	**0.85 (0.82-0.88)**	**0.28 (0.18-0.46)**	**0.89 (0.85-0.91)**	**0.47 (0.35-0.59)**
Yes	0.96 (0.94-0.97)		0.92 (0.89-0.93)		0.94 (0.92-0.95)	
No	0.59 (0.48-0.69)		0.35 (0.23-0.45)		0.49 (0.36-0.59)	
Classification	**0.64 (0.61-0.67)**	**0.37 (0.31-0.43)**	**0.73 (0.70-0.76)**	**0.15 (0.10-0.21)**	**0.55 (0.51-0.58)**	**0.33 (0.28-0.39)**
Cat. I /Normal	0.54 (0.39-0.64)		0.26 (0.18-0.33)		0.55 (0.48-0.62)	
Cat. II /Suspicious	0.67 (0.62-0.70)		0.83 (0.81-0.86)		0.42 (0.38-0.47)	
Cat. III /Pathological	0.63 (0.57-0.68)		0.26 (0.18-0.34)		0.66 (0.59-0.71)	

FHR, fetal heart rate; FIGO, International Federation of Gynecology and Obstetrics; ACOG, the American College of Obstetricians and Gynecologists, NICE, National Institute of Clinical Excellence.

Table 5. Interobserver agreement evaluated by the proportions of agreement (PA) with respective 95% confidence intervals (95% CI), for tracing classification by clinicians of the three study groups, according to their previous experience in cardiotocography interpretation.

	FIGO 1987	**ACOG 2010**	**NICE 2007**
	PA (95% CI)	PA (95% CI)	PA 95% CI
> 10 years experience			
Overall	**0.61 (0.56-0.67)**	**0.58 (0.53-0.63)**	**.54 (0.48-0.59)**
Cat. I/ Normal	0.52 (0.31-0.68)	0.31 (0.20-0.40)	.60 (0.52-0.68)
Cat. II/ Suspicious	0.67 (0.60-0.72)	0.70 (0.65-0.74)	.33 (0.25-0.41)
Cat. III/ Pathological	0.56 (0.47-0.64)	0.32 (0.13-0.48)	.66 (0.57-0.74)
6-10 years experience			
Overall	**0.62 (0.58-0.68)**	**0.73 (0.68-0.79)**	**.54 (0.49-0.60)**
Cat. I/ Normal	0.57 (0.38-0.71)	0.11 (0.03-0.19)	.48 (0.33-0.60)
Cat. II/ Suspicious	0.65 (0.59-0.71)	0.84 (0.80-0.88)	.44 (0.36-0.52)
Cat. III/ Pathological	0.60 (0.51-0.68)	0.12 (0.00-0.21)	.65 (0.58-0.72)
< 6 years experience			
Overall	**0.63 (0.58-0.69)**	**0.88 (0.85-0.92)**	**.55 (0.49-0.60)**
Cat. I/ Normal	0.52 (0.29-0.69)	0.46 (0.25-0.64)	.58 (0.48-0.68)
Cat. II/ Suspicious	0.63 (0.57-0.70)	0.94 (0.91-0.96)	.39 (0.32-0.46)
Cat. III/ Pathological	0.65 (0.58-0.73)	0.25 (0.00-0.53)	.67 (0.57-0.74)

FIGO, International Federation of Gynecology and Obstetrics; ACOG, the American College of Obstetricians and Gynecologists, NICE, National Institute of Clinical Excellence.

Table 6. Sensitivity and specificity of all tracings classified by observers as pathological or category III in prediction of newborn acidemia, by study group, and according to the number of years of experience of the clinicians in cardiotocography analysis (95% CI= 95% confidence intervals).

	Sensitivity % (95% CI)	**Specificity % (95% CI)**
FIGO 1987	**89 (52-98)**	**63 (55-71)**
> 10 years experience	90 (59-100)	67 (59-74)
6-10 years experience	86 (42-100)	64 (56-72)
< 6 years experience	90 (59-100)	58 (50-66)
ACOG 2010	**32 (10-67)**	**95 (90-98)**
> 10 years experience	38 (10-71)	92 (87-96)
6-10 years experience	24 (4-58)	95 (91-98)
< 6 years experience	33 (18-81)	98 (94-99)
NICE 2007	**97 (61-100)**	**66 (58-73)**
> 10 years experience	95 (56-100)	72 (64-79)
6-10 years experience	95 (59-100)	57 (49-65)
< 6 years experience	100 (59-100)	67 (59-75)

FIGO, International Federation of Gynecology and Obstetrics; ACOG, the American College of Obstetricians and Gynecologists, NICE, National Institute of Clinical Excellence.