Human factors affecting the interpretation of fetal heart rate tracings: an update

Susana Santo and Diogo Ayres-de-Campos

Purpose of review
Human factors can have an important impact on cardiotocography (CTG) interpretation and management decisions, and therefore may directly affect obstetrical outcomes.

Recent findings
It has been well demonstrated that there is wide observer disagreement over CTG interpretation, particularly in the evaluation of variability, decelerations, and overall tracing classification. The reasons behind this are still incompletely understood, but poor reproducibility can have a profound impact on the technology’s accuracy and on its efficacy. Some scientific societies have recently revised their guidelines for CTG interpretation, but no up-to-date universally accepted recommendation exists. In spite of some approximation between the major guideline sets, important differences still exist between them, and they remain complex and prone to memory decay. Regular training in CTG interpretation appears to result in increased knowledge, better observer agreement, and improved quality of care. Computer analysis has also been developed, but remains heavily dependent on staff to confirm interpretation and to decide clinical management.

Summary
An international consensus, comprising simpler and more objective interpretation guidelines, together with regular staff training, and improved decision support systems seem to be the way forward for this technology.

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

INTRODUCTION
Cardiotocography (CTG) was introduced into clinical practice during the 1960s for assessment of fetal well being during late pregnancy and labor. Expectations at the time were that it would lead to a decreased incidence of perinatal death and cerebral palsy. Reality has fallen very short of these expectations, perhaps not so surprisingly, as it was not known then that these outcomes are rarely caused by intrapartum hypoxia [1]. Nevertheless, it has been difficult to demonstrate that intrapartum CTG has any major impact on important obstetric indicators, with the exception of increasing operative delivery rates [2].

Despite these discouraging results, the technology remains widely used in industrialized countries [3,4]. Acquisition of tracings, their interpretation, and the resulting clinical decision continues to be almost entirely dependent on healthcare professionals, and this may constitute the weakest link of the technology. In this review we will focus on human factors that affect CTG interpretation, and identify possible ways of overcoming these limitations.

OBSERVER AGREEMENT ON CARDIOTOCOGRAPHY INTERPRETATION
It is well known that CTG interpretation is subject to high intraobserver and interobserver disagreement [5,6]. This disagreement involves both the identification of individual CTG features [7–9] and overall tracing classification [5,6]. Classification
of decelerations as early, variable or late [7–9], evaluation of variability [7,8] and overall classification of tracings as suspicious and pathological [6] seem to be the least reproducible aspects of CTG analysis. On the contrary, fetal heart rate (FHR) baseline estimation has been shown to be one of its most reproducible components [10].

More recently, observer agreement has been evaluated when CTG interpretation is performed according to the four-tier STAN (Neoventa Medical, Gothenburg, Sweden) guidelines and with the availability of electrocardiographic ST waveform data. The results suggest little or no improvements in the reproducibility of overall tracing classification. Ojala et al. [11] reported only a moderate interobserver agreement in overall CTG classification according to these guidelines and Westerhuis et al. [12] showed a higher agreement for normal and preterminal tracings than for intermediate or abnormal ones.

Confidential enquiries carried out in the United Kingdom have reported that substandard care related to the incorrect interpretation of intrapartum CTGs was present in over 50% of intrapartum related deaths [13]. Incorrect interpretation and/or inappropriate response to CTG changes have also been reported in large clinical series [14,15]. A recent study by Ayres-de-Campos et al. [16*] raises the possibility that these findings could, at least in part, be due to a more pessimistic evaluation conducted in the knowledge of an adverse neonatal outcome, in what is a poorly reproducible method. In this study, 100 consecutively acquired intrapartum tracings were evaluated twice by five experienced clinicians, without knowledge of the study’s objective or that they would be re-evaluating the same tracings. In the first round, no information on newborn outcome was provided. In the second round, carried out two months later, the order of the tracings was randomly altered and the newborn’s umbilical artery pH was indicated. In acidemic cases, this led to a significant increase in the number of tracings classified as pathological. The main differences were found in the evaluation of decelerations and variability. The study concludes that evaluators should not have access to neonatal outcome before analyzing tracings. In medical-legal cases, this requires the evaluation of several tracings, among which the index case is included.

Reproducibility is a basic characteristic of all diagnostic methods and one that has important consequences in their accuracy and efficiency. Costa-a-Santos et al. [17] showed that poor reproducibility can dramatically influence the results of validity studies, as well as those of randomized controlled trials. Moreover, there are important limitations to the mathematical assessment and interpretation of agreement measures that may account for some of the differences found between studies [18].

Little is known about the reason why CTG interpretation is poorly reproducible. Are healthcare professionals imprecise at following CTG guidelines or do they adapt them according to personal convictions? Are CTG guidelines too subjective in their definitions? Are they too complex to be understood? Are some or all of their concepts prone to memory decay over time? Are healthcare professionals not capable of using a systematic approach for identification of CTG features? Do they have eyeball limitations in this analysis? Are some healthcare professionals incapable of recognizing graphical patterns? Is the information acquired from the analysis of events too complex to synthesize into an overall classification?

Whatever underlining reasons there may be for the poor reproducibility of CTG interpretation, improving this aspect of the technology must be a major priority for all those who use it, lest it will always remain unable to reach its full potential. The different guidelines developed for the interpretation of CTG tracings remain at the heart of this problem, and will be considered in the following section.

**RECENT UPDATES IN CARDIOTOCOGRAPHY INTERPRETATION GUIDELINES**

Different CTG interpretation guidelines have been proposed since the early years of CTG. The first large
consensus efforts were carried out in the late 1970s by the American College of Obstetricians and Gynecologists (ACOG) and in the 1980s by the International Federation of Gynecology and Obstetrics (FIGO). The FIGO guidelines remain to date as the only worldwide consensus in this area, but no updated version has been published since the original document of 1987 [19]. The ACOG, on the contrary, has published several revised versions, the last of which in 2010, in association with the National Institute of Child Health and Human Development (NICHD) and the Society of Maternal-Fetal Medicine [20**]. The Royal College of Obstetricians and Gynaecologists (RCOG) has also published guidelines in 2001, and a revised version, in association with the National Institute of Clinical Excellence (NICE), was issued in 2007 [21].

A recent comparison between these three major sets of guidelines revealed some approximation between the definitions that they enclose [22**]. There is reasonable agreement on how to estimate the FHR baseline, identify accelerations and evaluate long-term variability. Normal baseline values, tachycardia, bradycardia, reduced variability, and the sinusoidal pattern also have relatively similar definitions. On the contrary, there are important disparities in the definitions of the different types of decelerations. Lack of objective descriptions of some FHR features and the interdependent nature of the definitions for baseline and periodic events are other important limitations of these guidelines.

All three guidelines now adopt a three-tier classification system and no major disagreements exist in the classification of tracings as normal. However, there is wide disagreement in the classifications of intermediate/suspicious and pathologic tracings.

The inexistence of universally accepted guidelines and the lack of objective CTG definitions have probably contributed to the limited effectiveness of CTG monitoring and are a major limitation to the recognition of this technology. Consensus guidelines are usually developed by experts, but they are applied by a large number of healthcare professionals with different levels of expertise and commitment to the area, who may experience difficulties in assimilation of concepts, application in individual cases, and retention of knowledge over time. This may have particular importance in medico-legal settings, where awareness and compliance with practice guidelines is frequently a central aspect of litigation [23].

Assimilation of concepts and retention of knowledge is likely to be easier when clear and simple associations exist between tracing classifications and clinical management, for example: normal tracings – continue monitoring, perform additional testing, or minor interventions to revert potentially hypoxic situations; pathological tracings – immediate intervention to revert causes of fetal hypoxia or prompt delivery if rapid reversal does not occur. Such an approach could also improve the reproducibility of clinical management based on the CTG, which has also been shown to be poor [6,12,24]. Some may argue that oversimplification may lead to rare cases of hypoxia being missed, and it is important to assure that errors fall more towards the side of over-intervention than to under-diagnosis. On the contrary, it must be kept in mind that many more errors probably occur because of the incomplete understanding and limited retention of what is a complex set of information. Keeping guidelines as simple as possible may be the key to generalized assimilation and recollection of concepts [22**].

**REGULAR TRAINING OF HEALTHCARE PROFESSIONALS**

The impact of training of health professionals on CTG interpretation is still incompletely understood. Several studies describing different educational strategies and training programmes have been published, together with an evaluation of their effect on learning objectives and on clinical outcomes. A recent systematic review of these studies suggests that training results in increased knowledge and interpretive skills, higher interobserver agreement, better management options, and improved quality of care [25*]. This review included studies that evaluated computer-based programmes, lectures, case studies, audits and simulated clinical scenarios. It concludes that regular CTG training programmes should be mandatory for all healthcare professionals. However, many questions remain unanswered, such as: which is the most efficient training method? Are they equally adequate for all professional groups? How often does training need to be repeated?

Combination of CTG training sessions with hands-on simulator-based courses has been shown, in retrospective observation studies, to be associated with a reduced incidence of adverse obstetrical outcomes [26], but a cause–effect relationship cannot safely be established and the effective components of this package remain unknown.

**SYSTEMS FOR COMPUTER ANALYSIS OF THE CARDIOTOCOGRAPHY**

Several computer systems for CTG analysis have been developed in the last decades, as a way to overcome the poor reproducibility of visual
interpretation. Some are limited to analysis of antepartum tracings whereas others also allow intrapartum analysis. A detailed review of these systems is beyond the scope of this article, but none of them currently recommend clinical management actions. The majority emit real-time alerts for changes detected in the FHR, relying on healthcare professionals to confirm tracing interpretation and to act accordingly. Therefore they are still heavily dependent on human factors.

A recent study reports that clinicians predict newborn umbilical artery pH more reproducibly and more accurately when they have access to computer analysis of tracings (that is estimation of baseline, evaluation of variability, identification of accelerations and decelerations and overall tracing classification) [27]. These findings suggest that healthcare professionals are consciously or unconsciously influenced by the results of computer analysis, perhaps leading to a more rigorous and systematic interpretation of the tracing, and consequently a better estimation of fetal oxygenation. However, the degree of influence of computer analysis on human interpretation may also be dependent on staff’s clinical experience and on their personal confidence with the system.

CONCLUSIONS

CTG remains heavily dependent on the judgment of healthcare professionals, and human factors affecting tracing interpretation can therefore have a profound effect on outcomes. It has been well demonstrated that CTG analysis is poorly reproducible, and this has been shown to have a profound impact on the method’s accuracy and efficacy. No universally accepted CTG interpretation guidelines currently exist, which has probably contributed to the limited effectiveness of the technology and is a major limitation for its recognition. International consensus in this area would be very welcome. Keeping guidelines as simple and as objective as possible could also be the key to a better assimilation of concepts and increased memory retention. Clear associations between tracing classifications and clinical management are also likely to improve the demonstration of efficiency. CTG training programmes appear to be an important tool to improve knowledge, observer agreement, and quality of care. Studies are needed to compare the effect of different training strategies and to define the best way to prevent knowledge decay. Decision support systems are under development and evaluation, but they still rely on human judgment to confirm interpretation and to decide management.

Acknowledgements

This work was partly financed by Eurostars project ESI649, combining funds from the EUREKA initiative and the European Commission.

Conflicts of interest

DAdE-C is one of the developers of the Omniview-SisPorto system (Speculum, Lisbon, Portugal) for computer analysis of CTG tracings, but receives no personal funding from its commercialization.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

Additional references related to this topic can also be found in the Current World Literature section in this issue (pp. 115–116).

The most recent guidelines for intrapartum FHR monitoring published by the American College of Obstetricians and Gynecologists, the National Institute of Child Health and Human Development and the Society of Maternal and Fetal Medicine. Clinical management in each classification category is considered.


A comparison between the FIGO, ACOG and RCOG/NICE guidelines for fetal monitoring. The main similarities and differences are pointed out and a simplified classification is proposed.


A systematic review of the impact of CTG training courses, including computer-based programmes, lectures, case studies, audits and simulated clinical scenarios.


When intrapartum tracings, monitored until very close to delivery, were randomly assigned to be presented with or without computer analysis to experienced clinicians, prediction of umbilical artery blood pH was significantly more reproducible and more accurate in the computer analysis group.