European Journal of Obstetrics & Gynecology and Reproductive Biology 205 (2016) 27–31

Contents lists available at ScienceDirect

European Journal of Obstetrics & Gynecology and Reproductive Biology

journal homepage: www.elsevier.com/locate/ejogrb

Interobserver agreement in CTG interpretation using the 2015 FIGO guidelines for intrapartum fetal monitoring☆

Mariana Rei[a,b,c,d,*], Sara Tavares[a,b], Pedro Pinto[b], Ana P. Machado[b], Sofia Monteiro[b], Antónia Costa[a,b,c,d], Cristina Costa-Santos[e,f], João Bernardes[a,b,c,d,f,g], Diogo Ayres-De-Campos[a,b,c,d,f]

[a] Department of Obstetrics and Gynecology, Medical School, University of Porto, Portugal
[b] S. Joao Hospital, Porto, Portugal
[c] Institute of Biomedical Engineering (INEB), Porto, Portugal
[d] Instituto de Investigação e Inovação em Saúde (i3s), Porto, Portugal
[e] Department of Health Informatics and Decision Sciences, Medical School, University of Porto, Portugal
[f] Centre for Research in Health Information Systems and Technologies (CINTESIS), Portugal
[g] Hospital Pedro Hispano, Unidade Local de Saúde de Matosinhos, Portugal

ARTICLE INFO

Article history:
Received 14 May 2016
Received in revised form 6 July 2016
Accepted 1 August 2016

Keywords:
Cardiotocography
Intrapartum fetal monitoring
Fetal heart rate
Reproducibility
Interobserver agreement
Reliability

ABSTRACT

Background: Visual analysis of cardiotocographic (CTG) tracings has been shown to be prone to poor intra- and interobserver agreement when several interpretation guidelines are used, and this may have an important impact on the technology's performance.
Objectives: The aim of this study was to evaluate agreement in CTG interpretation using the new 2015 FIGO guidelines on intrapartum fetal monitoring.
Study design: A pre-existing database of intrapartum CTG tracings was used to sequentially select 151 cases acquired with a fetal electrode, with duration exceeding 60 minutes, and signal loss less than 15%. These tracings were presented to six clinicians, three with more than 5 years' experience in the labor ward, and three with 5 or less years' experience. Observers were asked to evaluate tracings independently, to assess basic CTG features: baseline, variability, accelerations, decelerations, sinusoidal pattern, tachysystole, and to classify each tracing as normal, suspicious or pathologic, according to the 2015 FIGO guidelines on intrapartum fetal monitoring. Agreement between observers was evaluated using the proportions of agreement (Pa), with 95% confidence intervals (95%CI).
Results: A good interobserver agreement was found in the evaluation of most CTG features, but not bradycardia, reduced variability, saltatory pattern, absence of accelerations and absence of decelerations. For baseline classification Pa was 0.85 [0.82–0.90], for variability 0.82 [0.78–0.85], for accelerations 0.72 [0.68–0.75], for tachysystole 0.77 [0.74–0.81], for decelerations 0.92 [0.90–0.95], for variable decelerations 0.62 [0.58–0.65], for late decelerations 0.63 [0.59–0.66], for repetitive decelerations 0.73 [0.69–0.78], and for prolonged decelerations 0.81 [0.77–0.85]. For overall CTG classification, Pa were 0.60 [0.56–0.64], for classification as normal 0.67 [0.61–0.72], for suspicious 0.54 [0.48–0.60] and for pathologic 0.59 [0.51–0.66]. No differences in agreement according to the level of expertise were observed, except in the identification of accelerations, where it was better in the more experienced group.
Conclusions: A good interobserver agreement was found in evaluation of most CTG features and in overall tracing classification. Results were better than those reported in previous studies evaluating agreement in overall tracing classification. Observer experience did not appear to play a role in agreement.

Introduction

Since the introduction of cardiotocographic (CTG) monitoring in the 1960s, there have been variations in the way healthcare professionals interpret CTG tracings and decide subsequent clinical management. Different guidelines and poor intra- and interobserver agreement are important aspects of this technology, and

☆ Paper presentation information: Poster presented in the XXI FIGO World Congress of Gynecology and Obstetrics, Vancouver, Canada, 4–9th October 2015.

* Corresponding author at: Departamento de Obstetrícia e Ginecologia, Faculdade de Medicina da Universidade do Porto, Alameda Hernani Monteiro, 4200-319 Porto, Portugal.
E-mail addresses: marianacruzrei@gmail.com, marianarei@hotmail.com (M. Rei).

http://dx.doi.org/10.1016/j.ejogrb.2016.08.017
0301-2115/

they can have a significant impact on the results of validity studies, as well as on those of randomised controlled trials assessing effectiveness [1,2].

The first international consensus on CTG interpretation was promoted by the International Federation of Gynecology and Obstetrics (FIGO) in 1986 [3]. The resulting guidelines constitute an important landmark in the field, as they stimulated wide-scale international agreement on essential aspects of CTG monitoring, such as terminology, indications, acquisition techniques, and interpretation. However, with the passage of time they naturally became outdated. Several national scientific societies subsequently developed their own interpretation guidelines [4,5], resulting in the use of different terminology and interpretation criteria. A revised version of the FIGO guidelines was published in 2015, as a result of the largest international consensus effort promoted in the field [6].

The purpose of this study was to evaluate interobserver agreement in interpretation of CTG tracings using the new 2015 FIGO guidelines on intrapartum fetal monitoring.

Materials and methods

One-hundred-and-fifty-one CTG tracings were consecutively selected from a pre-existing database of intrapartum tracings, collected for research purposes in a tertiary care university hospital [7,8]. Patients were considered eligible for inclusion if they had 36 or more weeks of gestation, no known fetal malformations, a singleton fetus in cephalic presentation, were in the active phase of labor, and had an indication for continuous CTG monitoring. All patients were monitored until delivery. Selected cases were subsequently excluded if the tracing lasted less than 60 min, signal loss in the last hour exceeded 15%, an intrapartum event capable of interfering with fetal oxygenation occurred between tracing-end and delivery (difficult vaginal or abdominal fetal extraction, cord prolapse, maternal hypotension, shoulder dystocia, or anesthetic complications), or unreliable umbilical cord blood values were detected. All tracings presented to clinicians included the last 60 min acquired before delivery, at a paper speed of 1 cm/min, and no additional clinical information was supplied.

The 151 tracings were sent by email, together with the 2015 FIGO guidelines on intrapartum fetal monitoring, to six clinicians actively working in the labor ward of the same hospital, with different levels of expertise, and who volunteered to participate in the study. Three clinicians had more than 5 years of experience in the labor ward and three were residents with 5 or less years' experience. Observers were asked to evaluate the following basic CTG features: baseline (normal, tachycardia, bradycardia), variability (normal, reduced, increased) accelerations (present, absent), decelerations (present, absent; variable, late, repetitive and/or prolonged), sinusoidal pattern (present/absent) and tachysystole (present/absent). Observers were allowed to select the occurrence of more than one type of deceleration. They were finally requested to classify the tracing as normal, suspicious or pathologic.

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)" [19]. A Pa equal to 0.5 means that when an observer attributes a certain classification, there is a 50% probability that another observer will attribute the same classification. If the lower limit of the 95% CI for Pa was under 0.50, agreement was considered poor [9].

Reliability was evaluated with the kappa statistic (k-Light's kappa for *n* raters), measuring agreement beyond that expected by chance [kappa = (observed agreement − chance agreement)/(1 − chance agreement)]. Kappa 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 proportion of agreement and a low kappa when the prevalence of a given rating is very high or low [18]. The proportions of agreement relative to each individual category (proportion of specific agreement) help to understand that agreement is high in some categories and low in others [2,9,17,18]. 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 [10].

Table 1
Interobserver agreement and reliability for evaluation of basic CTG features and overall tracing classification, measured by the proportions of agreement (Pa) and kappa values, with respective 95% confidence intervals (95%CI).

	Number of times rated (906 ratings)	Pa [95%CI]	Kappa [95%CI]
Baseline		0.85 [0.82–0.90]	0.47 [0.35–0.58]
Normal	760	0.91 [0.89–0.94]	
Tachycardia	123	0.56 [0.44–0.65]	
Bradycardia	23	0.49 [0.07–0.74]	
Variability		0.82 [0.78–0.85]	0.30 [0.21, 0.39]
Normal	756	0.89 [0.86–0.92]	
Reduced	120	0.43 [0.32–0.52]	
Increased (saltatory)	17	0.20 [0.04–0.30]	
Accelerations		0.72 [0.68–0.75]	0.24 [0.15–0.33]
Yes	697	0.82 [0.78–0.84]	
No	208	0.38 [0.31–0.47]	
Decelerations		0.92 [0.90–0.95]	0.36 [0.19–0.52]
Yes	847	0.96 [0.94–0.97]	
No	58	0.42 [0.26–0.56]	
Tachysystole		0.77 [0.74–0.81]	0.42 [0.33–0.50]
Yes	234	0.56 [0.47–0.63]	
No	672	0.85 [0.81–0.87]	
Overall CTG classification		0.60 [0.56–0.64]	0.39 [0.33–0.45]
Normal	340	0.67 [0.61–0.72]	
Suspicious	358	0.54 [0.48–0.60]	
Pathological	208	0.59 [0.51–0.66]	