Sexual & Reproductive Healthcare 5 (2014) 195–198

Contents lists available at ScienceDirect

Sexual & Reproductive Healthcare

journal homepage: www.srhcjournal.org

# Single versus pairwise interpretation of cardiotochography, a comparative study from six Swedish delivery units

Hanna Åmark [a,*], Charlotte Millde-Luthander [a], Gunilla Ajne [b], Ulf Högberg [c], Hans Pettersson [a], Ingela Wiklund [d], Charlotta Grunewald [a]

[a] Karolinska Institute, Department of Clinical Science and Education, Unit of Obstetrics and Gynecology, Södersjukhuset, Stockholm, Sweden
[b] Department of Obstetrics and Gynecology, Karolinska University Hospital, Huddinge, Stockholm, Sweden
[c] Department of Women's and Children's Health, Akademiska Sjukhuset, Uppsala, Sweden
[d] Department of Clinical Sciences, Division of Obstetrics and Gynecology, Danderyd Hospital, Karolinska Institute, Stockholm, Sweden

ARTICLE INFO

*Article history:*
Received 28 August 2013
Received in revised form 15 May 2014
Accepted 19 May 2014

*Keywords:*
Cardiotocography
Fetal surveillance
Intrapartum
Pairwise interpretation
Cardiotocography interpretation

ABSTRACT

*Objective:* The aim of the study was to evaluate whether interpreting CTG pairwise brings about a higher level of correctly classified CTG recordings in a non-selected population of midwives and physicians.
*Study design:* A comparative study.
*Setting:* Five delivery units in Stockholm and one delivery unit in Uppsala, with 1589, 3740, 3908, 4539, 6438, and 7331 deliveries in 2011, respectively.
*Subjects:* 536 midwives and physicians classified one randomly selected CTG recording individually followed by a pairwise classification. The pairs consisted of two midwives (119 pairs) or one midwife and one physician (149 pairs), a total of 268 pairs.
*Main outcome measure:* The proportion of individually correctly classified CTG recordings versus the proportion of pairwise correctly classified CTG recordings.
*Results:* The proportion of individually correctly classified CTG's was 75% and the proportion of pairwise correctly classified CTG's was 80% (difference 5%, p = 0.12).
*Conclusions:* There was no statistically significant difference when CTG's were classified pairwise compared to individual classifications. The proportion of individually correctly classified CTG's was high (75%). There were differences in the proportion of correctly classified CTG recordings between the delivery units, indicating potential areas of improvement.



## Introduction

Fetal surveillance with cardiotochography (CTG) is commonly used as a diagnostic tool during labour although associated with difficulties in interpretation, which in some cases causes substandard care [1]. The clinical implication of the wide intra- and interrater variability in CTG interpretation is thus likely to lead to either unnecessary interventions or lack of appropriate management [2–5].

In order to meet the need for further education in fetal monitoring, a computer assisted learning programme for training and certification of cardiotochography (CTG) was developed (http://www.ctgutbildning.se). The programme was evaluated in a recent study in which 64% of 135 non-selected midwives and physicians correctly classified a CTG-recording before and 66% after the computer assisted learning programme (non-significant). Subsequently, some of the participants classified one more CTG recording together in pairs after passing the CTG programme. When analysing these pairwise CTG classifications, 79% pairs classified the CTG correctly. When the pairs were divided into two groups, i.e. midwife–midwife and midwife–physician, the proportions were 70% (16/23) and 85% (33/39), respectively (non-significant) [6]. Thus, even though no improvement in classifying CTG correctly after the completion of the programme was observed, the observation of the highest level of correctly classified CTG's when classified pairwise merits further exploration.

The purpose of this study was to further evaluate whether interpreting CTG pairwise brings about a higher level of correctly classified CTG recordings in a non-selected population of midwives and physicians and to evaluate if there is a difference between pairs with one physician and one midwife compared with pairs with two midwives.

*Abbreviation:* CTG, Cardiotocography

* Corresponding author. Tel.: +46739 89 4150.
*E-mail address:* hanna.amark@sodersjukhuset.se (H. Åmark).

http://dx.doi.org/10.1016/j.srhc.2014.05.004
1877-5756/

## Methods

### Setting

The study was carried out at five delivery units in the Stockholm region and Uppsala Akademiska Sjukhus, Sweden. The initial intention was to include all delivery units in the Stockholm region. However, one unit declined to participate, and to be able to include as many pairs as needed according to a tentative sample size calculation, we also invited Uppsala Akademiska Sjukhus. The delivery units included are of different sizes, as measured by numbers of deliveries in year 2011: n = 1589, n = 3740, n = 3908, n = 4539, n = 6438, n = 7331, and all of them handle both low and high risk deliveries.

### Materials

We used the same CTG pool as in the study evaluating the computer assisted learning programme [6]. We here give a brief description of the CTG pool, we refer to the original study for a more thorough account. Four experts, two midwives and two obstetricians, individually classified 55 intrapartum CTG recordings, received from a CTG database (Neoventa, Medical, Gothenburg, Sweden). The classification was done according to the Swedish modified version of the FIGO classification. According to the FIGO classification, each CTG belongs to one of the four categories normal, intermediary, pathological or preterminal. Among the 55 CTG recordings, there was an overrepresentation of intermediary and pathological CTG's. Earlier literature suggests that these categories are more difficult to correctly classify than the other two categories [2,3,7]. Of the 55 CTG's, 40 were classified with a 100% inter-individual agreement. These were considered reference standard and constituted the CTG pool in the original study and in our study. The distribution of categories was as follows: normal (n = 5), intermediary (n = 13), pathological (n = 17) and preterminal (n = 5).

There were 536 participants (midwives n = 387, physicians n = 149), all working at the six different delivery units. The participants were recruited at several random occasions at the units. Everyone who currently worked during these occasions, and had not yet participated, was invited into the study. Only 0.04% (22 out of 558) were invited and did not participate, which eliminates the risk of bias due to "self-selection". In total, 69% of the midwives and 68% of the physicians employed at the six different delivery units during the study period between May and September 2012 participated. Interpretation of CTG on a regular basis formed the prerequisite for participation.

### Design

Each participant was asked to individually classify a randomly selected CTG recording from the pool as normal, intermediary, pathological or pre-terminal. The participants were allowed to use their CTG-classification card, but they were not required or encouraged to use it. There was no time-limit. The participants were observed while they interpreted the CTG recording, to make sure that they complied with the given instructions. The classifications were done at different times of the day and during evenings whenever the participants worked. The profession (midwife or physician) of each participant was noted. Subsequently, each participant was paired with another participant at the same working shift. The pairing was done between those who were available at each shift by one of the authors, deliberately avoiding two physicians in the same pair, yielding 149 pairs consisting of one physician and one midwife, and 119 pairs consisting of two midwives. Each pair was asked to classify a randomly selected CTG recording from the pool. The two members of the pair were specifically instructed to cooperate in the classification. Apart from this, the instructions were the same as for the individual classifications.

### Statistical analysis

Before data were collected, a sample size calculation was carried out. We calculated the sample size required to achieve 80% power at 5% significance level when testing for equality between the proportion correctly individually classified CTG's and the proportion correctly pairwise classified CTG's. We assumed that the true proportions were equal to 65% (individually classified) and 80% (pairwise classified), respectively. We based the sample size calculation on the standard Fisher's exact test [8] (5% significance level, two sided). However, the calculation based on Fisher's exact test does not take the clustering of data into account. That is, the same individual appears twice in the data; once individually, and once as a member of a pair. In the actual analyses of data, we used Wald tests, which we adapted for clustered data. The required sample size was found to be 151 observations per group, i.e. 151 individually classified CTG's and 151 pairwise classified CTG's. The recruitment goal was eventually set to 302 pairs and a total of 604 participants, to be able to compare pairs consisting of two midwives and pairs consisting of one midwife and one physician, at the same power as the main comparison of individuals versus pairs. For the main comparison 151 pairs and a total of 302 participants were needed. The sample size calculation was done in Sample Power 2.0 [9].

We used Wald test (two sided, 5% significance level) and calculated 95% Wald confidence intervals to compare the proportions of correctly classified CTG recordings between the single versus pairwise classified CTG's [10]. The Wald test was also used for subgroup analysis between different delivery units and between pairs consisting of two midwives and pairs with a physician and a midwife. To take the clustering of data into account, robust ("sandwich") standard errors were used in the denominator of the test statistics and in the limits of the confidence intervals. The statistical analyses were carried out in Stata [11].

To investigate whether our results may suffer from bias due to a possible imbalance between the proportion of pathological and intermediate recordings in the group of individual classifications and the group of pairwise classifications, we carried out a comparison of individual versus pairwise classification stratified on CTG type.

### Ethical considerations

The participants were orally informed and data management was anonymous.

## Results

Table 1 displays the proportion of individual and pairwise correctly classified CTG recordings for each separate delivery unit, and for the whole sample. The proportion of correctly classified CTG's

**Table 1**
Individual versus pairwise classification of CTG.

| Delivery unit | Correct individual classification | Correct pairwise classification | Difference | P-value | Confidence interval (95%) |
|---|---|---|---|---|---|
| A n = 82 | 0.68 | 0.68 | 0.00 | 1.0 | –0.18, 0.18 |
| B n = 96 | 0.71 | 0.81 | 0.10 | 0.11 | –0.03, 0.24 |
| C n = 100 | 0.74 | 0.90 | 0.16 | 0.01 | 0.04, 0.28 |
| D n = 166 | 0.75 | 0.80 | 0.05 | 0.3 | –0.05, 0.15 |
| E n = 70 | 0.93 | 0.77 | –0.16 | 0.06 | –0.32, 0.04 |
| F n = 22 | 0.76 | 0.82 | 0.06 | 0.7 | –0.28, 0.39 |
| **Total n = 536** | **0.75** | **0.80** | **0.05** | **0.12** | **–0.01, 0.10** |

n = number of individuals, number of pairs is n/2.