33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Computerized intrapartum electronic fetal monitoring: analysis of the decision to deliver for fetal distress

Antoniya Georgieva, Stephen J Payne, Mary Moulden, and Christopher WG Redman

Abstract— **We applied computerized methods to assess the Electronic Fetal Monitoring (EFM) in labor. We analyzed retrospectively the last hour of EFM for 1,370 babies, delivered by emergency Cesarean sections before the onset of pushing (data collected at the John Radcliffe Hospital, Oxford, UK). There were two cohorts according to the reason for intervention: (a) fetal distress, n_1 = 524 and (b) failure to progress and/or malpresentation, n_2 = 846.**

The cohorts were compared in terms of classical EFM features (baseline, decelerations, variability and accelerations), computed by a dedicated Oxford system for automated analysis – OxSys. In addition, OxSys was employed to simulate current clinical guidelines for the classification of fetal monitoring, i.e. providing in real time a three-tier grading system of the EFM (normal, indeterminate, or abnormal).

The computerized features and the simulated guidelines corresponded well to the clinical management and to the actual labor outcome (measured by umbilical arterial pH).

I. Introduction

THIS study continues the development of OxSys, a clinical decision support system that uses computerized analysis of the fetal heart rate in labor, recorded by Electronic Fetal Monitoring (EFM). EFM is routinely used in labor as a cardiotocogram (Fig. 1) that is visually assessed by clinicians.

The fetal response to the stress of labor is reflected in the changes of the fetal heart rate pattern [1]. It is often difficult to interpret the EFM but failure to intervene when the fetus is hypoxic can cause severe acidemia (and possible neurological damage and neonatal morbidity). Conversely, misjudgment can cause unnecessary Caesarean birth [1,2]. It has been established that there is poor inter- and intra-observer agreement between clinical experts when they interpret the intrapartum EFM [3]. In addition, experts make many false positive diagnoses because of the mediocre predictive power of EFM for fetal acidemia [1-4]. Moreover, there are no simple, linear relations between the EFM features and fetal acidity [1, 6, 7]. In this context, the potential of computerized analyses of the EFM has been widely recognized [8]. It is objective and allows large datasets to be studied.

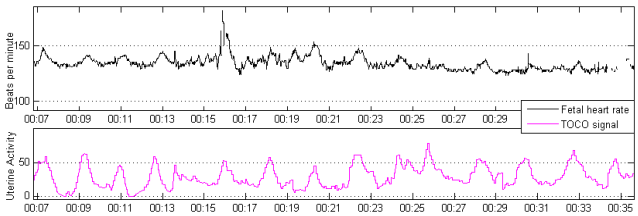

Fig. 1. Example of a 30 minute electronic fetal monitoring (EFM).

There are several clinical guidelines for intrapartum EFM interpretation and management although they still contain contradictory aspects [5]. They aim to promote objectivity and can improve labor outcomes after clinical staff are trained in their use [2]. Nevertheless, computerization and data-driven optimization of clinical guidelines has been delayed, because adverse events are rare (<0.05%) and very large digital archives are necessary for statistically meaningful results. Schiermeier et al. [7] demonstrated on a small study (370 deliveries) that fetal pH measured from scalp blood samplings can be related to computerized clinical guidelines. The authors suggested further work to produce improved computerized algorithms or guidelines.

In this preliminary work, we describe the use of OxSys to computerize current clinical guidelines. We selected study cases who were delivered before the second stage of labor by Cesarean section for fetal distress or other unrelated reasons. The aim was to compare these cohorts in terms of computerized EFM features and computerized interpretation of current guidelines. Such investigations are necessary to learn what factors positively or negatively influence the clinical decision making and provide insights how to improve this by computerized interpretation (the Cesarean section for fetal distress is a common clinical event which is under-studied [9]).

II. Data Selection and Computerized Features

A. Data Selection

Selected were 1, 370 cases, delivered in Oxford, UK, by emergency Cesarean sections during the first stage of labor, i.e. before the onset of pushing (Fig. 2). Cases were included only if the reason for intervention was specifically documented as (a) fetal distress (524 cases) or (b) failure to progress or malpresentation (846 cases). Cesareans for other reasons were excluded here because it was not absolutely clear that there was no aspect of fetal distress involved. The diagnosis of fetal distress depended on EFM.

A. Georgieva is with the Nuffield Department of Obstetrics and Gynaecology and with the Institute of Biomedical Engineering, University of Oxford (corresponding author: tel. +44 (0) 1865 857 852; e-mail: antoniya.georgieva@obs-gyn.ox.ac.uk).

S. J. Payne is with the Institute of Biomedical Engineering, University of Oxford (e-mail: stephen.payne@keble.ox.ac.uk).

M. Moulden is with the Nuffield Department of Obstetrics and Gynaecology, University of Oxford (e-mail: mary.moulden@obs-gyn.ox.ac.uk).

CWG Redman is with the Nuffield Department of Obstetrics and Gynaecology, University of Oxford (e-mail: christopher.redman@obs-gyn.ox.ac.uk).

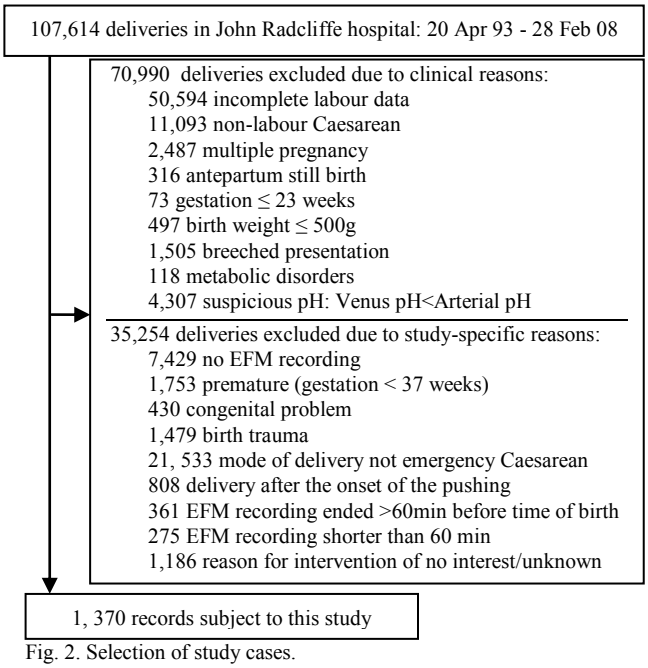

Fig. 2. Selection of study cases.

Another criterion of inclusion was the availability of reliable umbilical arterial cord blood measurements so that acidemia at birth could be objectively diagnosed. A pH threshold of 7.15 was used to discriminate acidotic (low arterial pH) from non-acidotic (normal arterial pH) babies. A relatively high threshold [10] was appropriate because (i) the cases did not undergo second labor stage which would have lead to even lower pH values, and (ii) the aim was to identify the interventions that have successfully prevented even more severe acidemia. Table I shows the predictive rate of the clinical interventions for the chosen pH threshold.

Finally, only cases with at least 60 min of continuous EFM, ending within 60 min of birth were considered. Both groups had similar intervals between the time of birth and the end of EFM. This interval was less than 10 min in over 50% of cases.

TABLE I
CLASSIFICATION OF THE INTERVENTIONS (FETAL DISTRESS OR OTHER REASON), MEASURED BY UMBILICAL ARTERIAL PH.

TP	*FP*	*PPV*	*Mis-classified*	*Kappa*
114	410	21.8%		
FN	*TN*	*NPV*		
39	807	95.4%	32.8%	0.2
Sensitivity	*Specificity*	*Total*		
74.5%	66.3%	1,370		

True Positive (TP) – Intervention for fetal distress & pH≤7.15;
False Positives (FP) – Intervention for fetal distress & pH>7.15;
False Negative (FN) – Intervention for other reason & pH≤7.15;
True Negatives (TN) – Intervention for other reason & pH>7.15.

B. OxSys Feature Extraction

OxSys has been developed and optimized [11-13] to extract standard EFM features: baseline, variability, decelerations (late, variable or early), accelerations and contractions. It also extracts non-standard features such as approximate entropy and kurtosis, but these are not considered here. Table II summarizes which features were included. All records were pre-processed as previously described [13].

The features are calculated in a 15 minute moving window (moving step of 5 minutes), updating the values every 5 minutes. Hence, in the last 60 minutes there were ten 15-minute windows, from which the OxSys features were extracted. The median of the ten values for each feature was taken.

TABLE II
HOW OXSYS FEATURES WERE DERIVED

Feature	*Algorithm Summary*
Baseline	Calculated using morphological filters [11].
Deceleration & Acceleration	Calculated using morphological filters and heuristic thresholds [13].
Short Term Variability*	Mean absolute difference of neighbouring heart rate measurements. Moderate range thresholds were estimated as the 5th and 95th centiles respectively on a set of healthy babies.
Long Term Variability*	Mean difference of the highest peak and the lowest trough in subsequent 1 minute windows. Moderate range thresholds were estimated as the 5th and 95th centiles respectively on a set of healthy babies.

* Accelerations and decelerations are excluded from the signal before calculating Short and Long Term Variability.

C. Computerized simulation of clinical guidelines

OxSys was configured to simulate the latest clinical guidelines published by the American College of Obstetricians and Gynecologists (ACOG) [14]. These offer a three-tier system to classify EFM segments of arbitrary length at any time in labor. Category 1 segments are classified as normal and include all of the following: baseline between 110 and 160bpm, moderate variability, and no late or variable decelerations. Category 3 segments are abnormal and include either (*i*) absent variability with bradycardia or absent variability with recurrent late/variable decelerations or (*ii*) a sinusoidal pattern. Segments that are neither Category 1 nor 3 are an indeterminate Category 2.

We used OxSys to apply these criteria and assign an ACOG Category to each of the ten 15 min windows of interest. It must be kept in mind that:

- The ACOG guidelines are designed for visual EFM interpretation and are not specified for precise automation.
- OxSys calculates both short term variability and long term variability and both were included with equal weight in the simulation.
- The thresholds for moderate and absent variability differ between the OxSys and the visual ACOG criteria. We use data driven thresholds to define moderate (and thus reduced or increased) variability (Table II).

- Currently OxSys does not detect sinusoidal rhythms but this will be added in due course. The sinusoidal is a very rare and ominous pattern.

For these reasons, the OxSys simulation of the above three-tier classification system cannot be exact. The findings reported below (Section III) should be considered only in this context.

III. Results and Discussion

The derivation of OxSys features is summarized in Table II. Acidosis at birth was used to classify the accuracy of the clinical decisions to intervene. Thus an acidotic baby, delivered for fetal distress, was a true positive while a non acidotic baby delivered for the same reason was a false positive. An acidotic baby delivered for failure to progress was a false negative while a non-acidotic baby delivered for the same reason was a true negative. The individual features and their groupings into guideline categories were analyzed by acidotic (pH≤7.15) and non-acidotic (pH>7.15) outcomes, by reason for intervention and by true or false positives and negatives (Fig. 3).

Fig. 3 shows that there is a significantly higher number of decelerations and significantly lower number of accelerations in the acidotic group as compared to the non-acidotic, although the differences are not large (around half a deceleration or acceleration). This mean difference is even larger (one deceleration and one acceleration respectively) when the two cohorts defined by reason for intervention are compared. This is encouraging confirmation that OxSys deceleration and acceleration are related to both the fetal state and the clinical decision making (decelerations are known to indicate fetal distress and accelerations – fetal health). However, the average baseline and variability (short or long term) are similar across the first four groups in Fig. 3.

The 524 cases delivered for fetal distress were broadly similar to the acidotic group (153 cases) in terms of the computerized features and guideline categories suggesting that the clinical decisions were on average correct. Of particular interest is the fact that there were no significant differences in the computerized EFM and guideline features between the true positives and the false positives. We conclude that none of the considered computerized features can aid their separation. However, the false negatives (39 cases) showed a raised baseline and a significantly reduced variability when compared to the true negatives. Hence there may be a small number of acidotic cases, undetected by clinicians, with computerized features that are different from those of the other groups (including true positives). This false negative group merits further investigation to improve the detection of compromised cases by computerized analysis.

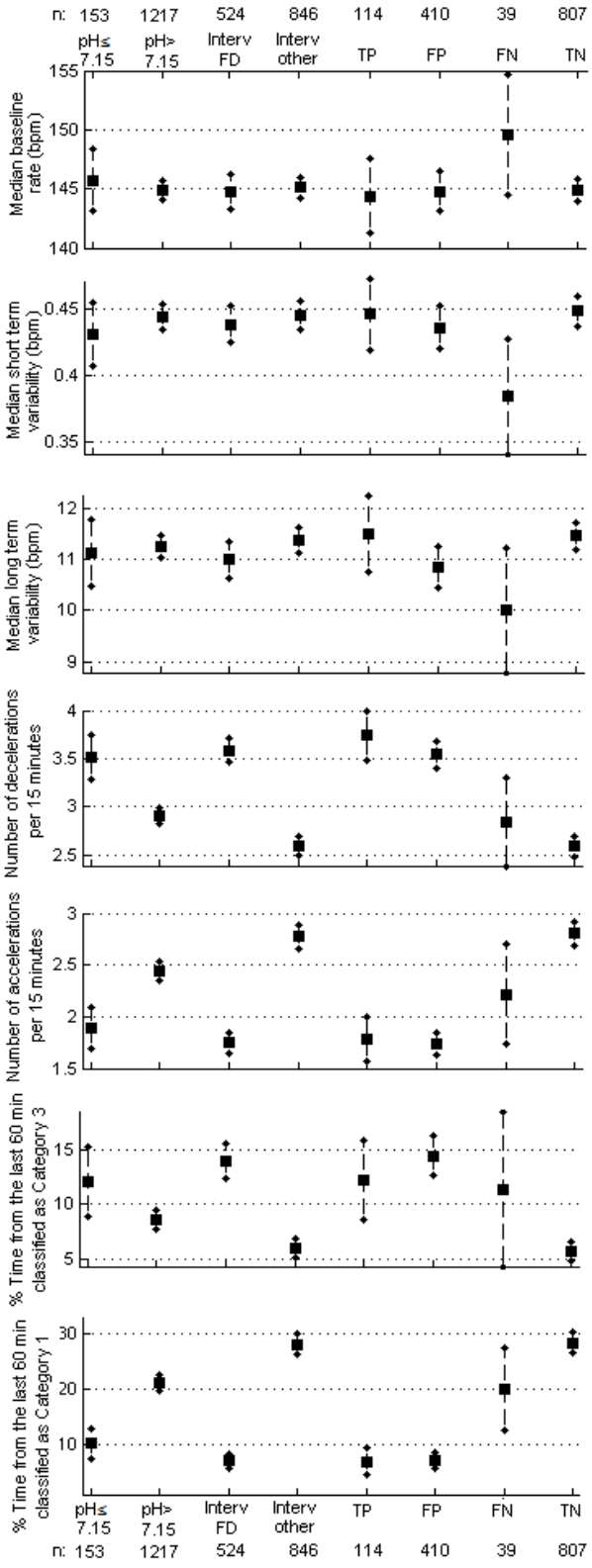

Fig. 3. Mean value (95% confidence interval) of the computerized features stratified in relevant groups of interest: acidotic (arterial pH≤7.15) or else (pH>7.15); deliveries for the specific reason of fetal distress (Interv FD) or else (Interv other); true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN), as defined in Table I.

We also gain an interesting insight into the relationship between clinical diagnoses using EFM and the ACOG guidelines which were used (which are broadly similar to the UK guidelines [5]). The average proportion of time (in the last 60 min) where the EFM was classified as Category 1 (normal) is three times shorter in the group delivered for fetal distress (less than 10%) than in the group delivered for other reason (nearly 30%). The converse pattern was observed for Category 3 (abnormal). This good correspondence between the computerized guidelines and clinical decisions, is further illustrated in Fig. 4. It shows the cumulative distribution of cases with respect to percentage of time the EFM was classified as normal or abnormal. About 70% of the Cesareans for fetal distress did not have a single episode of normal (Category 1) trace in the last 60min of EFM (as opposed to 32% in the group of interventions for other reason). Similarly, 72% of the interventions for other reasons did not have a single episode of abnormal (Category 3) monitoring. This concordance is despite the fact that the bedside diagnoses used clinical information (epidural, oxytocin use, parity, gestation, and so on) not available to the computerized analyses.

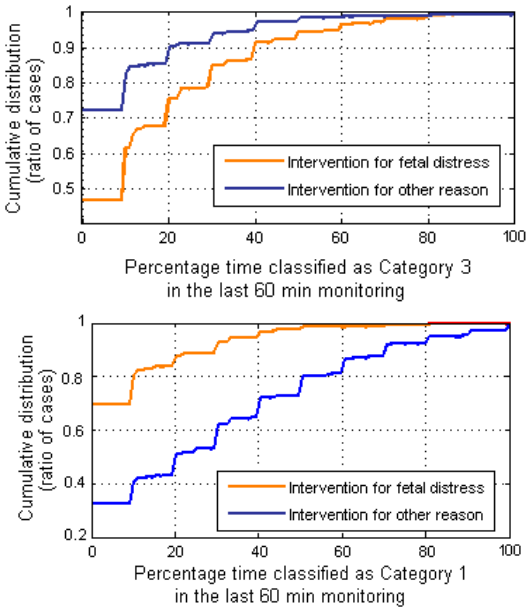

Fig. 4. Relation between the computerized simulation of clinical guidelines [14] and the actual decision that has been made (Category 1 is classified as normal and Category 3 as abnormal).

IV. Conclusion

We studied by computerized methods the clinical decision making in 1,370 labors that ended with emergency Cesarean deliveries. Delivery for the reason of fetal distress was associated with increased number of decelerations, reduced number of accelerations, more abnormal segments and fewer normal ones (as defined by the simulated guidelines). We concluded that the computerized EFM features emulated well the clinical assessments. Moreover, these can be successfully used to simulate contemporary clinical guidelines for EFM interpretation and labor management.

The next step will be to optimize the computerized guidelines over data from many deliveries and develop objective data-driven decision support system in labor.

References

[1] L. Bennet and J. Gunn, "Fetal responses to asphyxia", in Fetal and neonatal brain injury, 4th ed. D.K. Stevensonet al., Eds. Cambridge UP, 2009, pp. 143-62.

[2] T. Draycott, T. Sibanda, L. Owen, V. Akande, C. Winter, S. Reading, and A. Whitelaw, "Does training in obstetric emergencies improve neonatal outcome?," *BJOG*, vol. 113, pp. 177–82,2006.

[3] S.P. Chauhan, C.K. Klauser, T.C. Woodring, "Intrapartum nonreassuring fetal heart rate tracing and prediction of adverse outcomes: interobserver variability," *AJOG*, vol. 199, pp. 623.e1-623.e5, 2008.

[4] D.A. Grimes, J.F. Peipert, "Electronic fetal monitoring as a public health screening program: the arithmetic of failure," *Obstet Gynecol*, vol. 116, pp. 1397-1400, 2010.

[5] D. Ayres-de-Campos D and J. Bernardes, "Twenty-five years after the FIGO guidelines for the use of fetal monitoring: Time for a simplified approach", *Int J Gynecol Obstet*, vol. 110, pp. 1-6, 2010.

[6] L.C. Pello, S.K. Rosevear, G.S. Dawes, M. Moulden, and CWG Redman, "Computerized fetal heart rate analysis in labor," *Obstet Gynecol*, vol. 78, pp. 602-10, 1991.

[7] S. Schiermeier, S. Pildner von Steinburg, A. Thieme, et al., "Sensitivity and specificity of intrapartum computerised FIGO criteria for cardiotocography and fetal scalp pH during labour: multicentre, observational study," *BJOG*, vol. 115, pp. 1557-63, 2008.

[8] J. Westgate, "Computerizing the cardiotocogram (CTG)", in *Medical Informatics in Obstetrics and* Gynecology, D. Parry and E. Parry, Eds. Medical Info Science Reference, 2009, 151-58.

[9] K.K. Roy, J. Baruah, S. Kumar, et al., "Cesarean section for suspected fetal distress, continuous fetal heart monitoring and decision to delivery time," *Indian J Pediat*, vol. 75, pp. 1249-52, 2008.

[10] M. Alberry, S. Fuente, P.W. Soothill, "Prediction of asphyxia with fetal gas analysis", in *Fetal and Neonatal Neurology and Neurosurgery*, 4th ed. M.I. Levene, and F.A. Chervenak, Eds, Churchill Livingstone, 2009, pp. 528-41.

[11] S. Cazares, "Automated identification of abnormal patterns in the intrapartum," Ph.D. dissert., Dept. Eng. Sci., Oxford Univ., 2002.

[12] A. Georgieva, S.J. Payne, C.W.G. Redman, "Computerised electronic fetal heart rate monitoring in labour: automated contraction identification." *Medical & Biological Engineering & Computing*, vol. 47, pp. 1315-20, 2009.

[13] A. Georgieva, S.J. Payne, M. Moulden, C.W.G. Redman, "Automated fetal heart rate analysis in labor: decelerations and overshoots," in *Proc. 36th Int Conf on Application of Mathematics in Engineering and Economics*, Sozopol, Bulgaria, 2010.

[14] American College of Obstetricians and Gynecologists, "Practice bulletin 106: Intrapartum fetal heart rate monitoring: nomenclature, interpretation, and general management principles," *Obstet Gynecol,* vol. 114, pp. 192–202, 2009.