Computer Assisted Devices for Intrapartum Fetal Heart Rate Monitoring

Study Designs and Patient Populations

Deborah A. Wing, MD
Professor and Director
Division of Maternal-Fetal Medicine
Department of Obstetrics-Gynecology
University of California, Irvine
UCSF Antepartum and Intrapartum Management Course
June 11, 2010

Current Limitations

• Outcome measure not related to FHR monitoring patterns
• Lack of standardized interpretation of FHR patterns
  - Leads to poor interobserver and intraobserver consistency
• Disagreement re: algorithms for intervention for specific FHR patterns
• Inability to demonstrate the reliability, validity, and ability to FHR monitoring to allow for timely obstetrical intervention

Evidence-base

• 1985: Dublin trial demonstrated no benefit
  - Long-term follow-up of those infant with seizures found no permanent neurologic deficits
  - Infants with cerebral palsy did not necessarily have abnormal FHR patterns
• 1995: Thacker meta-analysis (n>18,000)
  - Frequency of neonatal seizures 0.8% with continuous EFM versus 1.1% with intermittent auscultation (RR 0.5)

History of FHR Monitoring

1950’s: Concept and product design
1960’s: First commercially available model
  - Test for asphyxia
  - Recognition of early asphyxia so that timely obstetrical intervention could avoid asphyxia-induced brain damage or death
1980’s: Market permeation

Recognition of early asphyxia so that timely obstetrical intervention could avoid asphyxia-induced brain damage or death
WHAT HAPPENED IN THE US?
• American independence (or ingenuity?)
• All professional organizations, inc. ACOG and AWOHNN, state that some form of intrapartum monitoring is necessary
  • The unspoken message: Cost-savings

WERE THE EXPECTATIONS FOR EFM TOO HIGH?
• Cerebral palsy due to intrapartum asphyxia accounts for only 10% of cases
• Cannot prevent all catastrophic asphyxial injuries
• Cerebral palsy is rare

WHAT’S ACHIEVABLE?
• Umbilical artery pH ≤ 7.0
  • Incidence 3/1000
  • Predictor of neurologic and other organ damage
  • If threshold is umbilical artery pH < 7.1, then frequency as high as 2.5%

WHAT’S THE PROBLEM IN INTERPRETATION?
• Overemphasis on late deceleration complex
  • Two mechanisms of development
• Overstated value of umbilical cord integrity
• De-emphasis of FHR variability
• Poor understanding of fetal physiology, as well as the dynamics and evolution of FHR patterns
• Inability to define cause-effect relationship
• Lack of standard nomenclature
**COMPUTER-ASSISTED DEVICES**

- Computerized pattern recognition algorithms used in conjunction with conventional monitoring and diagnostic modalities for clinical decision support
- Purpose: Reduce errors
- Can influence detection, description, diagnosis or reporting tasks

**COMPUTER-ASSISTED DEVICES: HOW DO THEY WORK?**

- CAD algorithms first trained to identify patterns using data from a finite sample of patients with or without abnormalities, as a “training set”
- Once trained or optimized, the CAD system can be used on new patients to detect similar abnormal patterns and discriminate them from normal ones
  - Needs to be sufficiently sensitive and specific to aid clinicians in their identification and/or assessment of abnormalities
  - Application to a new data set or a “test set” that is independent of the training set

**COMPUTER-ASSISTED DEVICES**

- **Type 1**: Detects minimum/maximum thresholds for simple EFM monitoring (e.g., FHR >150 bpm, FHR <120 bpm; FHR variability <5 bpm, etc.)
- **Type 2**: Detects one or more specific EFM patterns that are generally recognized of clinical interest (late deceleration, hyperstimulation, sinusoidal FHR patterns, etc.)
- **Type 3**: Risk stratification for future event of clinical concerns (e.g., biphasic event or urgent delivery)

**COMPUTER-ASSISTED DEVICES: HOW DO THEY WORK?**

- Signal or image processing
  - Smoothing, sharpening, histogram equalization, etc.
- Segmentation
  - Temporal boundaries mapped to important time intervals (epochs) or key clinical notations
- Feature calculation
- Classification/discrimination
- Thresholding and output
  - Warnings or monitoring alarms
  - Mark on data section (imaging)
• General information
  • What is it targeted to detect?
  • For whom is the device intended?
  • On what kind of system(s) will the CAD be installed?

• Intended Device Usage
  • Will operations and settings be manual (by the physician), semi-automatic, or completely automatic (nothing is under physician control)?
  • What kind of output is generated by the CAD device? Is there user feedback?
  • What is the reading mode?

• Which classification methods are used in the CAD algorithm (e.g., simple threshold, decision tree, linear classifier, neural network)?
  • How were they optimized?

• Training and Test Databases
  • Which types of abnormal and normal patterns and what types of patients are represented in the training and test sets?
  • Are the training and test sets representative of the intended use population?
  • What are the sample sizes of the training and test sets?

• Processing
  • What processing is performed on the data (e.g., filtering, noise reduction, normalization)?

• Features
  • Which features were computed or evaluated during algorithm development?
  • How were features selected?
  • What is the mathematical formulation of the feature?
  • How does the feature relate to medicine/biology?

• Algorithm stability
  • Is the CAD algorithm performance robust to minor changes to the algorithm or perturbations to the training or test data?
  • Note, the stability of an algorithm increases as the number of training cases increases, the number or dimensionality of initial features decreases, or the complexity of the CAD decreases herbal
COMPUTER-ASSISTED DEVICES: HIERARCHICAL MODEL OF EFFICACY

- **Level 1**: Technical efficacy
- **Level 2**: Diagnostic accuracy
- **Level 3**: Diagnostic thinking

<table>
<thead>
<tr>
<th>Level</th>
<th>Technical Efficacy</th>
<th>Therapeutic Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Physical performance measurements, preclinical stand-alone and bench tests</td>
<td>Effect of diagnostic imaging or test on therapeutic management of patients</td>
</tr>
<tr>
<td>2</td>
<td>Diagnostic accuracy</td>
<td>Patient outcome</td>
</tr>
<tr>
<td>3</td>
<td>Diagnostic thinking</td>
<td>Societal efficacy</td>
</tr>
</tbody>
</table>

- **Level 4**: Therapeutic efficacy
- **Level 5**: Patient outcome
- **Level 6**: Societal efficacy

- **Expected value of test information in terms of gains on QALYs; also cost per QALY gain**
- **Cost-effectiveness and/or cost-benefit analysis from societal viewpoint**

STUDY DESIGNS

- **Clinical**
  - Prospective randomized, controlled trial
  - Prospective, non-randomized
  - Retrospective Multi-reader Multi-Case Study (MRMC)

- **Non-Clinical**
  - Stand-alone performance testing
  - Reproducibility testing

RCT

- Randomization to two or more reading modalities such as with/without CAD
  - Can be stratified by reader
- Regarded as best trial design to measure effect of CAD on clinical endpoints such as adverse event rate and intervention rate
RCT: LIMITATIONS

- For rare events such as intrapartum fetal death or neonatal encephalopathy, sample size must be large for study to have sufficient power
  - Proxies such as umbilical arterial pH <7.10 and Bdefc>12 mmol/L could be used
- Ethical considerations
- Hawthorne effect

RCT: MODULATIONS

- Masked CAD readings with post-hoc analysis on performance

INFANT TRIAL: INTELLIGENT SYSTEM SUPPORT DECISION MAKING IN LABOR MANAGEMENT USING CTG

- Recently launched by MHS
- Objectives: 1) To determine whether intelligent decision-support can improve interpretation of CTG and thereby improve labor management for women requiring continuous EFM
  - A) Identify more clinically significant heart rate abnormalities
  - B) Result in more prompt and timely action on clinically significant heart rate abnormalities
  - C) Result in fewer “poor neonatal outcomes”
  - D) Change incidence of operative interventions

INFANT TRIAL: OBJECTIVES

2) To assess whether use of intelligent decision-support improves quality of routine care received by women undergoing continuous EFM during labor.
   - Intended to determine if decision-support software decreases risk of suboptimal intrapartum care
3) To determine whether use of decision-support software is cost-effective in terms of incremental cost per poor perinatal outcome prevented
4) To determine whether use of decision-support software has any effect on longer term neurodevelopment of children born to women participating in INFANT trial
**INFANT Trial: Study Design**

- RCT of 46,000 women
- Individual randomization of women who are judged to require continuous EFM based on NICE criteria
- Singleton or twin pregnancy
- ≥35 weeks' gestational age
- No known gross fetal abnormality
- K2MS Guardian system

Protocol provided courtesy of Peter Brocklehurst, MD, Oxford Perinatal Epidemiology Unit

**NICE GUIDELINES: INDICATIONS FOR CONTINUOUS EFM**

- Significant meconium-stained liquor
- Abnormal FHR detected by intermittent auscultation (<110 bpm; >160 bpm; any decelerations after contraction)
- Maternal pyrexia (defined as 38.0°C once or 37.5°C on two occasions 2 hours apart)
- Fresh bleeding developing in labor
- Oxytocin augmentation
- Patient’s request

**INFANT TRIAL: STUDY DESIGN**

- K2MS/Guardian system will provide alerts regarding eligibility and can randomly allocate subjects to “no decision-support” or “decision support”
- All de-identified information collected by K2MS/Guardian system within participating center can be accessed centrally

**INFANT TRIAL: STATUS**

- MHS directed single site six-month pilot
- As of April 15, 2010: n=185
  - Low refusal rate (32%)
  - Anticipated to begin enrollment in all centers in July 2010
- [http://www.npeu.ox.ac.uk/infant/](http://www.npeu.ox.ac.uk/infant/)
GUARDIAN/K2MS SYSTEM

Development occurred along two main paths: Expert system and methods for feature extraction from the CTG

Condition to be an expert:
1. Clinician must be experienced and still practicing obstetrics
2. Clinician should be regarded as an “expert” by their peers and superiors

K2MS/GUARDIAN SYSTEM

Evaluation and validation: 5-step process

1. Evaluation of the expert system
2. Evaluation of the integrated system
3. External validation of the system
4. In-house on-line trial
5. Multicenter randomized trial

GUARDIAN/K2MS SYSTEM

System integration

• Developed individual modules using conventional numerical algorithms to measure and classify baseline heart rate, accelerations, heart rate variability, timing of decelerations and signal quality.
• Algorithms were designed to be flexible to allow experts to shape classification criteria and have been validated.
• Feature extraction modules, together with 5th generation neural network, were then combined with expert system to form complete system capable of examining cases off-line.
• Outputs of feature extraction methods are presented graphically to provide assessment of their functioning.
• Functionality of system was then visually examined by expert to confirm that feature extraction methods and expert system seemed reasonable.

Fetal Monitoring Training System

http://www.k2ms.com/products/fmts.html
INFANT TRIAL: STUDY DESIGN
- Teaching system (90% permeation across UK)
- K2MS Guardian System adopts rule-based approach so that clinician can always ask for explanation as to why a course of management is recommended. In this way, informed judgment is possible.
- Expert system interprets combined data using a database of over 400 rules which are used to recommend action.

INFANT TRIAL: SHORT-TERM OUTCOME MEASURES
“Poor neonatal outcome” composite to include:
a) all deaths (intrapartum stillbirths plus neonatal deaths, i.e., deaths up to 28 days after birth) except deaths due to congenital anomalies,
b) significant morbidity: neonatal encephalopathy (mild, moderate and severe); other admission to NICU within 48 hours of birth for >48 hours which are associated with adverse intrapartum events including respiratory symptoms and seizures.

INFANT TRIAL: LONG-TERM OUTCOMES
Developmental Quotient (DQ) at two years of age
Acknowledgement that sample size is unlikely to be sufficient to determine difference between groups; estimate of intrapartum asphyxia in mature infants leading to moderate or severe cerebral palsy 1.5 – 2.5/1000.

INFANT TRIAL: SECONDARY OUTCOMES
- Cord-artery pH <7.05 and Bdef >12 mmol/l
- Distribution of cord-artery pH and base deficit
- Distribution of fetal blood sample data
- Apgar score <4 at 5 minutes
- Intrapartum stillbirth
- Neonatal death
- Seizures
- Admissions to neonatal unit within 48 hours of birth for at least 48 hours with evidence of:
  - Encephalopathy
  - Feeding difficulties
  - Respiratory illness
INFANT TRIAL: PROCESS OUTCOMES

- Number of CTG abnormalities identified in two arms
- Time taken to respond to alerts and alarms by undertaking fetal blood sampling or prompt delivery

REQUIRES USE OF PROSPECTIVE METHOD OF DATA COLLECTION AND RETROSPECTIVE EVALUATION OF DIGITIZED DATA TO DETERMINE WHEN ALERT OR ALARM SHOULD HAVE OCCURRED

- Number of routine measurements recorded during labor, inc. number of vaginal examinations, frequency of maternal temperature, HR, B/P, epidural analgesia, and labor augmentation

INFANT TRIAL: INCIDENCE OF PRIMARY OUTCOME AND EFFECT SIZE

- Poor neonatal outcome: 50% reduction, from 3 to 1.5 per 1000 with α=0.05, β=0.10
- Developmental quotient: 1% difference in mean DQ with α=0.05, β=0.10. Standardized mean of 100 with SD of 12.8. Sample size: 7000 babies
- Severe metabolic acidosis: 28% RR reduction in incidence with over 80% power in those babies who have cord artery pH measured

INFANT TRIAL: SECONDARY INFANT OUTCOMES

- In subset of 7000 surviving children followed-up at age 2 years:
  - Developmental Quotient components
  - Cerebral palsy
  - Late deaths (after neonatal period)

INFANT TRIAL: Assumptions of baseline incidences

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor perinatal outcome composite</td>
<td>3 per 1000 births</td>
</tr>
<tr>
<td>Intrapartum stillbirth</td>
<td>0.35 per 1000 births</td>
</tr>
<tr>
<td>Neonatal death</td>
<td>0.70 per 1000 births</td>
</tr>
<tr>
<td>Severe and moderate neonatal encephalopathy</td>
<td>1.3 per 1000 births</td>
</tr>
<tr>
<td>Combined outcomes</td>
<td>1.7 per 1000 (95% CI 1.5-1.9; range 0.8-2.3; 18 hospitals, Trent, 2003-2004)</td>
</tr>
<tr>
<td></td>
<td>1.9 per 1000 (95% CI 1.6-2.3; range 0.6-2.3; 12 hospitals, Yorkshire Neonatal Network, 2004-2005)</td>
</tr>
</tbody>
</table>
INFANT TRIAL: SECONDARY MATERNAL OUTCOMES

- Cesarean section
- Any operative intervention for either fetal distress and failure to progress
- Any episode of fetal blood sampling
- Length of first stage, length of active second stage, total length of labor
- Episiotomy

MRMC MODEL: 17 EXPERTS

Objectives: To investigate 1) whether intelligent computer system would compare with experts using CTGs, clinical information and fetal blood sampling, and 2) whether experts could be consistent and agree in their management of labor

Design: 50 cases reviewed by each expert and system on two occasions, ≥1 month apart. CTGs scored in 15-minute epochs, and clinical information provided when requested

Outcomes: Consistency and agreement in recorded scores, agreement and timing for cesarean deliveries, FBS sampling rates, intervention in cases with poor outcome and intervention in cases with good clinical outcome

Results: The system:
1. Agreed with experts and better than chance (67.3%, kappa=0.31, p<0.001)
2. Was highly consistent (99.16%, kappa=0.98, p<0.001) when used by two operators independently
3. Recommended no unnecessary intervention in cases with normal delivery and good condition (UA pH >7.15, UV pH >7.20, 5-min Apgar ≥9 and no resuscitation)
4. Recommended cesarean in 11 cases
5. Identified as many of the birth asphyxiated cases (UA pH <7.05 and Bdecf >12, and 5-min Apgar ≤7 with neonatal morbidity) as the majority of experts and one more than was acted upon clinically

MRMC MODEL: 17 EXPERTS

Results:
1. Experts were also found to be consistent and to agree.
2. Good agreement in the cases and timing of cesarean recommendations
3. Most experts did not recommend operative intervention in cases with normal delivery and good outcome, but did recommend operative interventions in 10 of 12 cases delivery with UA pH <7.05.
4. In one cases of birth asphyxia, 14 of 17 experts and the system failed to recommend intervention
CLINICAL TESTING: CAVEATS

- Skewed outcome data with type 3 CADs using contemporary databases
  - If clinician intervenes to deliver baby, then future monitoring is lost and the false positive rate is driven upward. In this circumstance, the only data to be collected without bias is NPV

- Lack of blinding could skew results in modified RCT where outcome is reduction in adverse event and reduction in cesarean rate