Re-conceptualizing fetal monitoring

Mark I. Evans[1,2,3], Robert D. Eden[1], David W. Britt[1], Shara M. Evans[1,4], Barry S. Schifrin[1]

[1] Fetal Medicine Foundation of America; [2] Comprehensive Genetics, PLLC; [3] Department of Obstetrics & Gynecology Mt. Sinai School of Medicine, New York, NY USA; [4] Department of Maternal and Child Health, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, NC USA

ABSTRACT

Electronic fetal monitoring (EFM) has been used extensively in labor for over 40 years despite its appreciated failure to identify, in a timely fashion, and help prevent a large proportion of cases of neonatal encephalopathy and cerebral palsy. Our analysis suggests that the poor performance of EFM derives from a fundamental misunderstanding of the differences between screening and diagnostic tests, large inter-observer variability in its interpretation as a result of very subjective classifications, failure to follow the physiology of fetal compromise, and poor statistical modeling for its use as a screening test. We have recently developed a new methodology, the fetal reserve index (FRI) which contextualizes the interpretation of EFM by 1. breaking EFM down into four components: heart rate, variability, accelerations, and decelerations; and then 2. adding increased uterine activity, and 3. risk factors (maternal, fetal, and obstetrical) to create an 8-point algorithm. In a direct comparison of the ACOG monograph criteria, ACOG Category system and the FRI, the FRI performed much better in identifying cases at risk before damage had occurred, and reduced both the need for emergency deliveries and overall Cesarean delivery rates.

KEYWORDS

Fetal reserve index, electronic fetal monitoring, cerebral palsy, neonatal encephalopathy, technology assessment.

Basic principles

Most clinicians are not aware that there exists a distinct field of technology innovation and assessment which is generalizable across disciplines and individual situations. It has its own literature, societies and meetings, [1,2] as well as norms, models and expectations. Understanding this can avoid pitfalls commonly seen in the introduction of new techniques — pitfalls that, with a grasp of the history of previous approaches, could have been completely predicted and often prevented. First there is the phase of "development". As a generalization, new ideas have often originated from academic settings, where ideas are conceptualized, tested, possibly patented and published [1,2]. Demand is created, and then it starts to move out into practice. As demand for a new concept expands, the originators cannot handle the demand, others want "in" on the game, and new utilizers emerge. This is the "diffusion" phase. It is well understood that, during diffusion, utilization rapidly expands, but complications often skyrocket. This is where misinterpretation of electronic fetal monitoring (EFM) fits in. These same concepts equally apply to medical therapies, surgical procedures and laboratory tests [1,2].

Overall, the incorporation of new technologies in medicine has proceeded at a slower pace than seen in many industries. The culture of medicine, while desperately seeking new approaches to critical problems, has also been, simultaneously, notoriously resistant to radical changes. The timing of adoption of new techniques is often very variable in practice with physicians/institutions/countries ranging, across a spectrum, from being "early adopters" to "late adopters". There are many components underlying how such variability comes into play, including technological capabilities, the resources available for bringing in new technologies, the cost/benefits of such developments, return on investment, and perceived liability reductions and exposures deriving from such moves [1,2].

Article history
Received 17 Jan 2019 - Accepted 25 Mar 2019

Contact
Mark I Evans; evans@compregen.com
131 E 65th St. New York, NY 10065, USA - Tel. +1 212 288 1422

Overall, a move from one technology to its replacement requires two constructs: 1. that the new technology seems ready and is an improvement over the existing one, and 2. that providers become uncomfortable with staying with the old approach. With minimal exceptions, there is never universal agreement that the new technology should immediately replace the old one — just as there is not usually universal acceptance that a new paradigm should replace an older one. It is not a simple, mechanistic process [3]. Occasionally, the abandonment of an old practice (e.g. the use of diethylstilbestrol for prevention of miscarriage) comes rapidly because of overwhelming evidence [4]. External and political forces can also come into play. In the 1980s, testing blood samples for HIV was developed in France and gained acceptance there. Even after its implementation in France and in other countries, in the United States, such testing remained prohibited by the Food and Drug Administration. Then, all of a sudden, they announced that this testing was to go from prohibited to mandatory, essentially overnight. Needless to say, there was considerable chaos as laboratories and blood banks had to reverse course with little notice [5].

On the positive side, in the case of myocardial infarction,

the identification of the role of CPK isoenzymes changed the paradigm of diagnosis of that condition from a clinical gestalt to a "lab test" fairly rapidly [6]. Acceptance, then, rests on a combination of factors that must be in place for the process to move forward. It may depend as much on the ability to resolve practical problems and hold promise for the future as on technical assessment of evidence. To put all this in context, currently debated evolutions include the use of cell-free fetal DNA (cffDNA) versus procedures with microarrays, and panethnic Mendelian screening [6-9]; cffDNA clearly identifies an increased percentage of Down syndrome, but it comes at the cost of abandonment of diagnostic procedures in which microarray analysis could detect a far higher number of serious disorders [7,8]. Conversely, pan-ethnic Mendelian screening is still underutilized. Even in well identified risk groups such as the Ashkenazi Jewish population, such screening finds more abnormalities that are not within the typical Ashkenazi panel than are within it [9]. One possible influence on the divergence of utilization of cffDNA versus microarrays and expanded Mendelian screening has been that cffDNA was rushed into practice with high marketing budget pressures by companies with only minimal refereed publications, whereas microarrays followed a much longer, traditionally rigorous process of multiple studies, including an NIH sponsored multicenter randomized trial, before entering practice [7-10]. Similarly, EFM was adopted quickly and before many basic principles had been established and properly understood [11]. There are 7 criteria that it is generally felt necessary to consider before deciding to screen for a condition (Table 1) [12]. Not all tests currently in use, however, actually follow these guidelines, which can lead to disproportionate expectations, expenditures and complications from follow-up testing that was likely unnecessary. Furthermore, the goal of a screening program is to detect the maximum number of "affected" individuals for the least number found to be screen positive. Where to put the cutoff point is, by definition, arbitrary, but it must be maintained to ensure maximum efficiency. Specifically, a program is not judged by whether any particular patient's problem can be identified [12].

Medicine, and particularly obstetrics, involves repetitive use of both diagnostic and screening tests. Most patients and many physicians do not understand the difference. Diagnostic tests are meant to give a definitive answer, may carry risks, may be expensive, and are only meant for patients at a risk high enough to warrant their use. Conversely, screening tests are meant for "everyone", and all they are asked to do is divide individuals with high enough risk to warrant diagnostic testing from those without such risk. They do not give definitive answers [12]. How well they do their job is defined by the metrics of sensitivity, specificity and positive and negative predictive value. These principles of evaluation, or performance characteristics, were introduced into practice in the 1970s by Galen and Gambino [13], and they establish the boundaries of a playing field and a scoring system within which competitors, which may possibly offer better ways of doing things, can be evaluated. Sensitivity and specificity are relatively well-known test properties. Less often used are the criteria that summarize some of the critical tradeoffs that clinicians face. One such tradeoff is the relative number of true positive cases and false positive cases, which is reflected in the ratio of the two, i.e. true positives/false positives (TP/FP), and expressed as the positive likelihood ratio (PLR). A second tradeoff involves the relative number of false negatives and true negatives, reflected in the ratio of the two, i.e. false negatives/true negatives (FN/TN), and expressed as the negative likelihood ratio (NLR). Competitive approaches should be characterized by high PLRs and fractional NLRs.

Table 1 Criteria for screening programs.

CRITERIA FOR SCREENING PROGRAMS
Relatively frequent disorder
Impairing or fatal
Beneficial intervention possible
Good performance metrics (high sensitivity, specificity and predictive values)
Prompt testing and follow up
Benefits outweigh costs
Voluntary and Educational

Applications to electronic fetal monitoring

Physiological basis of EFM

In 2008, the American College of Obstetricians and Gynecologists (ACOG) introduced a three-tiered "category system" (CAT system) based on the presumed presence of fetal acidemia [14]. Category I (CAT I) represents a completely reassuring tracing (i.e. absent acidemia). Category III (CAT III) suggests imminent danger (or presence of injury) and the need for immediate delivery from presumed acidemia to prevent or decrease worsening of the fetal injury. Category II (CAT II) shows "elements of concern", but it is "intermediate" (meaning non-diagnostic). There is no specific understanding of or agreement on how hypoxia or acidosis came to be present, or how much time the fetus has left before irreversible neurological injury occurs. Furthermore, and more concerning, is that there is no recommended course of action other than "continued observation". Implicit is the assumption that, without acidemia sufficient to cause neurological injury (an "essential" parameter of intrapartum injury), the fetus is otherwise "normal". The CAT II tracing has received considerable criticism and been redefined by others, but such reformulations have not successfully improved outcomes [15-17]. As per the principles articulated above, the goal of a screening program is to identify cases at high risk with enough discriminatory power to signal concern but to do so before irreversible sequelae occur [12]. Only then can EFM be a true screening test for neurological injury accompanied by the opportunity to correct the pathophysiology before irreversible fetal neurological injury occurs.

There is no obvious pathophysiological basis for the ACOG's three-tiered system in fetal heart rate (FHR) pattern surveillance. Based on the retrospective analysis of the ACOG monographs on neonatal encephalopathy and cerebral palsy (CP), the CAT system can actually only serve as a diagnostic screening test for injury that has already occurred or is in the process of occurring [18, 19]. By the time the CAT III stage is reached, it is often already too late to effectively alter the process of fetal injury, even with emergency operative delivery

(EOD). The EFM characteristics of CAT III tracings (which coincide with the acid-base values of the ACOG Monograph criteria) are: absent or sinusoidal variability of the FHR baseline, absence of FHR accelerations, FHR decelerations with late recovery, absent variability during the recovery and tachycardia (often >180 bpm), or an agonal baseline. Unfortunately, logic requires that in order to prevent neurological injury from occurring, the CAT III diagnostic criteria must be replaced by criteria that, being applicable earlier in the pathophysiology, serve to screen for the risk of neurological injury if labor were to continue without resuscitation or intervention. Therefore, it is imperative to understand the relationships among the EFM monitoring parameters since the presence of one abnormality affects the others in an associative relationship. The parameters do not exist in a vacuum as the onset of fetal hypoxia/acidosis triggers a cascade of physiological neurological changes which do not occur simultaneously. We have analyzed patients who entered labor with CAT I tracings and delivered a baby with CP without another apparent cause other than labor [20, 21]. We analyzed the degree of abnormality of the individual EFM parameters, and the timing and duration of abnormalities during the course of labor and delivery. Fetuses normal at the onset of labor that went on to develop CP demonstrate a characteristic pattern: hypoxia/ischemia, and predictable deterioration to the point of injury in association with excessive uterine activity (≥8 uterine contractions/20 minutes) (Table 2). The apparent ontogeny of hypoxia/asphyxia in pregnancies where fetuses are "normal/uninjured" at the onset of labor starts with the occurrence of contractions. For control patients (good outcomes), the average length of labor was 11.3 hours. For those who developed CP, it was 17.7 hours. There were several other differences in the average time to initial appearance of EFM abnormalities and the order of deterioration of EFM variables (Table 3), as there was progressive and relatively orderly loss of reassuring characteristics of EFM parameters. With traditional overall assessment of the FHR tracing, we noted, as internal benchmarks for our studies, both the point when the fetus became "no longer reassuring (which we define as Point A)" and then the point at which it became "injured (which we define as Point B)." While almost all the CP cases reached both Points A and B, only 30% of CP cases reached CAT III, and when this did occur it occurred later than Point B in every case, and most often in the 2nd stage of labor, within 20 minutes of delivery. These aspects have been discussed more extensively in our previous publications [16, 22-24]. Only by correlating the pathophysiological relationships between the onset of hypoxia/asphyxia and the pattern of deterioration of the EFM parameters can an effective EFM "screening" protocol be created. In its simplest terms, the analysis of decelerations is based on an assessment of their impact on baseline rate and variability. This was first revealed in the 1970s in the normal outcomes of fetuses with reactive positive oxytocin challenge test (OCT) results (i.e., late decelerations associated with accelerations and normal variability) and the frequent adverse outcomes in patients with non-reactive positive OCT results (i.e., late decelerations and diminished variability, tachycardia, etc.) [25]. The decelerations seen under these circumstances frequently represented fetal breathing movements and not significant asphyxia [26]. These same principles exist today. The deterioration of reassuring EFM parameters should be used to determine caution, and prompt intrauterine resuscitation (IR), and intervention when necessary, rather than waiting for the presence of a CAT III tracing and irreversible fetal neurological injury. As shown by our previous data, in the normal fetus, in the presence of contractions, reduction of oxygen availability due to impaired uterine, umbilical or cerebral blood flow begins with decelerations well before any alteration in the baseline features in response to uterine contractions. Thus, by the time variability disappears, the fetus has already spent considerable time and effort compensating for impaired oxygen availability / blood flow. To require, as the CAT system does, complete absence of variability before the pattern can be called CAT III ignores the general ontogeny of these changes which is: 1. FHR decelerations with decreasing, but not yet absent, FHR variability, 2. mild elevation of baseline rate with slow return to baseline following contractions, 3. Loss of FHR accelerations, and finally 4. fetal tachycardia (>160 bpm) or bradycardia (<110 bpm). It is a well understood axiom in medicine that one cannot treat something until it has been diagnosed. Unless the emerging changes in fetal tracings are recognized before neurological injury occurs, there are no options for earlier intervention and prevention of neurological injury. Thus, we believe that a screening method must ask how the fetus is able to tolerate essentially each contraction, from the point of view of how much "reserve" it has to withstand the next one, etc. And these two aspects must be assessed in a contemporaneous fashion.

Table 2 Electronic fetal monitoring variables.

	Reassuring	Non-Reassuring (Point A)	Abnormal (Point B)
Uterine contractions	≤ 8/20 Minutes	>8/20 Mins	>12/20 MINS
FHR baseline variability	5-25 BPM	<5 or ≥15 BPM	0 BPM or >25 BPM, Sinusoidal
FHR accelerations	>15 X 15 BPM/15 Secs	<15 BPM/15 Secs	<10 BPM/15 SECS
FHR decelerations	No late return to baseline	Late return to baseline (i.e. +OCT)	Late/Prolonged decelerations
Baseline FHR (BPM)	110-160 BPM	>15 BPM Rise since admission (<160)	<110/>160 BPM, Agonal

*FHR: Fetal heart rate; **BPM:** beats per minute*

Table 3 Deterioration of electronic fetal monitoring variables during labor in cerebral palsy cases (hours to reach).

	Non-Reassuring (Point A)	Abnormal (Point B)	Delivery
Order of appearance			
Increased uterine activity	7.7 Hours	10.4 Hours	11.8 Hours
Abnormal FHR Variability	6.1 Hours	8.6 Hours	10.1 Hours
Late return to baseline	4.8 Hours	7.6 Hours	8.7 Hours
Non reactivity	3.7 Hours	6.2 Hours	7.7 Hours
Abnormal FHR baseline	0.30 Hours	3.1 Hours	4.5 Hours

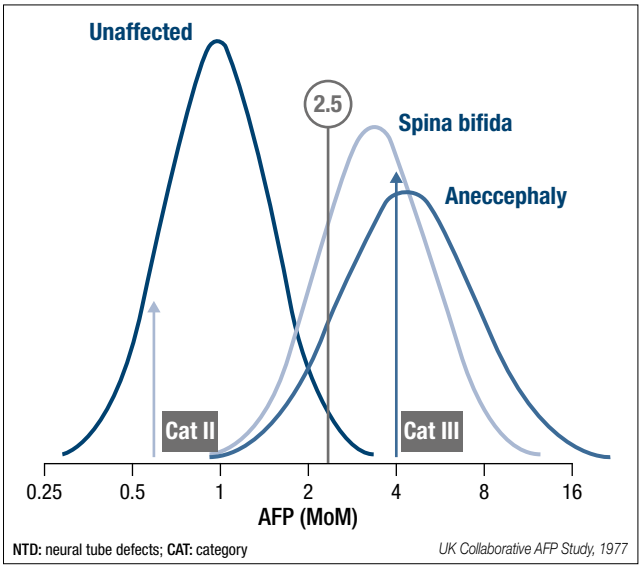
FHR: Fetal Heart rate

Statistical failures of the cat system

EFM parameters must be re-set to clearly identify (yes/no) fetuses at the precise point in time that they first suggest that the fetus is having to compensate (successfully) for diminished oxygen availability and/or impaired blood flow, instead of waiting until signs of decompensation and acidemia are present (or "diagnostic" as in the CAT III criteria). To serve as a successful screening test, the parameters must be identified as being abnormal when the characteristics have only reached "no longer reassuring" as opposed to already identified as "injured."

The CAT system is a clear screening system failure because almost 80% of the patients reach CAT II, and CAT III comes far too late in the pathophysiology. By contrast, maternal serum alpha fetoprotein screening for neural tube defects in the 1970s had a cutoff point at 2.5 multiples of the mean (MoM) that identified about 90% of affected cases for a false positive rate of 5%.27 CAT III would be equivalent to moving the aforementioned cutoff to 4.00 MoM (Figure 1). At that far right point on the distribution curve, the positive predictive value of screen-positive cases would be very high, but the false negatives would be pervasive. As mentioned, CAT III is reached too late in the pathophysiology, such that interventions would be unlikely to prevent neurological injury. The positive predictive value is very high only because many such babies have already suffered damage. The very high proportion of false negatives suggests that CAT III is incapable of preventing neurological injury. Alternatively, the CAT II cutoff would be about 0.7 MoM, whose sensitivity would be very high only because it includes so much of the population. However, with a 70-80% false positive rate, it violates the fundamental principles of screening tests mentioned earlier, rendering it clinically useless for prioritizing management towards a manageable number of screen positives for whom intervention can be successful. Further to this point, the development of the many proposed, complicated management protocols requiring sophisticated and nuanced interpretations by highly skilled practitioners to manage CAT II EFM is prima facie proof of the validity of our concerns and the failure of the CAT system as screening test and management framework [16]. Indeed, a majority of babies who develop neonatal encephalopathy and CP from the events of labor never have a CAT III tracing appreciated; they are mostly only CAT II. With either cutoff point, the performance is statistically substandard. Clinically, CAT was an improvement upon then existing approaches, but in retrospect, it had significant opportunity for further improvement.

Figure 1 CAT II and CAT III superimposed upon maternal serum alpha fetoprotein distributions for controls and neural tube defect (NTD) pregnancies. In the 1970s, MSAFP cutoff of 2.5 multiples of the median (MoM) detected 90% of NTDs for 5% false positives. CAT III would be equivalent of having cutoff of 4.0 MoM. Achieves very high positive predictive value but poor sensitivity. CAT II with cutoff about 0.7 MoM would have high sensitivity because false positive rate would approach 75%. Original figure courtesy of Dr. Howard Cuckle.



NTD: neural tube defects; **CAT:** category

UK Collaborative AFP Study, 1977

Developing a new approach

We have developed a modified approach to the interpretation of EFM [20-24, 28]. Our risk scoring system formally includes both antepartum and intrapartum risk factors that contribute to adverse neurological outcomes in newborns. Our conceptual notion is that interpretation of FHR should be optimized not for the recognition of asphyxia, but for the prevention of injury and for avoidance, by conservative measures, of the need to "rescue". We defined a new term, the "fetal reserve index" (FRI), which is a weighted calculation of various maternal, obstetrical and fetal risk (MOFR) factors along with quantitative component FHR interpretation and the presence of increased uterine activity (IUA) (Table 4) [20-24, 28]. The FRI categorizes the various risk factors on the basis of their anticipated effect on maternal well-being, placental and cerebral perfusion, and the probability of safe vaginal delivery. All definitions used are standard as per ACOG criteria, except that we define IUA as ≥5 contractions per 10 minutes rather than 6. We have explained these aspects in detail elsewhere [20-24, 28]. The FRI was initially calculated for each 20-minute segment of monitoring. In the calculation, each of 8 categories is assigned a score of "1" if the

Table 4 Components of the fetal reserve index.

COMPONENTS OF THE FETAL RESERVE INDEX
FHR
Baseline variability
Accelerations
Decelerations
Increased uterine activity
Maternal risk factors
Obstetrical risk factors
Fetal risk factors
Each factor scored as 1 if normal, and 0 if not.
Maximum of 8/8 = 100%.
Green zone: >50%
Yellow zone: 50% to 26%
Red zone: ≤25%

category is deemed normal and "0" if it is considered abnormal (Table 4). The MOFR variables are static, that is, once point reductions in each category occur, then they remain until the fetus is delivered. The EFM and IUA variables, however, are dynamic and therefore may change as the characteristics of the FHR tracing change often in response to: 1. the clinical onset of labor complications and progression to the second stage of labor, and 2. the onset of pushing and descent of the fetal head in the lower pelvis. The FRI was calculated for the number of points divided by 8 and multiplied by 100 to give a percentage. All 8 categories being normal would result in an FRI of 100 (8/8). Loss of points would result in FRI values of: 87.5 (7/8), 75.0 (6/8), 62.5 (5/8), 50.0 (4/8), 37.5 (3/8), 25.0 (2/8), 12.5 (1/8), and 0 (0/8). For clinical simplification, the scores were then divided into 3 zones: green >50%, yellow 50% to 26%, and red ≤25%. An abnormal FRI was defined as ≤25 (corresponding to the "red zone"). Entering the red zone is not to be taken as a call for immediate delivery, but rather as a cause for immediate attention by senior staff, who can evaluate the situation. IR efforts should usually be the first course of action, such as: stopping oxytocin, repositioning the patient, increasing IV fluids, and administration of oxygen by mask. As a good analogy, reaching the red zone can be likened to defending a corner kick, as opposed to a penalty kick. Most of the time it will turn out fine. Entering the red zone should also start a "shot clock" (as in basketball), and our management protocol is to allow up to 40 minutes to get out of the red zone. Failure to do so would start a 30 minute to delivery protocol, as per the ACOG guidelines with, as an overall resulting, delivery CP occurring usually within 1 hr of turning red which is less time than any of our CP cases were in the Red zone [20-24, 28]. To assess the performance of the FRI, our first study was a direct comparison of the postnatal ACOG monograph criteria, CAT III criteria and FRI in a data set of 60 singleton term babies who developed CP — all of whom had entered labor with CAT I tracings [15]. For none of them, even in retrospect, were there other apparent causes of their neurological compromise beyond labor issues. These infants were compared to 200 controls with normal outcomes.

In this study, the Apgar scores of the CP cases were much lower, as were their pH measurements which averaged 7.03. However, only a third 27% of the cases had a pH of <7.00, arguing against the rigidity of the ACOG monograph criteria, which required 7.00 for labor-induced issues to be considered possible [21]. The pattern of the FRI showed substantially lower scores for CP babies than controls. Only 22% of controls reached the red zone, and they were there for an average of 1 hour. CP babies "turned red" earlier in labor and stayed in the red zone for an average of over 5 hours (Figures 2, 3).

Figure 2 Representative control cases. Each column is a different patient. Time proceeds downward in 20 minute intervals.
Case A: 27 y.o. multigravida at 39 weeks with asthma. Following AROM, onset of decreased variability, loss of accelerations, FRI entered the red zone after onset of variable contractions with delayed recovery and IUA. IR performed with discontinuation of oxytocin as FRI reverted back to the yellow zone. Normal spontaneous vaginal delivery (NSVD) of 3650 g, Apgar's 8/9 with pH of 7.23.
Case B: 23 y.o. multigravida at 35 weeks with obesity and oligohydramnios. The yellow zone was entered after an epidural given following AROM. Variable decelerations developed with late recovery in 2nd stage. NSVD of 2690 g, Apgar's 8/9, pH 7.30.
Case C: 19 y.o. multigravida at 40 weeks with SROM. Experienced onset of IUA for 4 hours resulting in decreased variability, prolonged FHR deceleration, and loss of accelerations. Entered the red zone but IR was performed. NSVD of 3720g, Apgar's 8/9, pH 7.22.

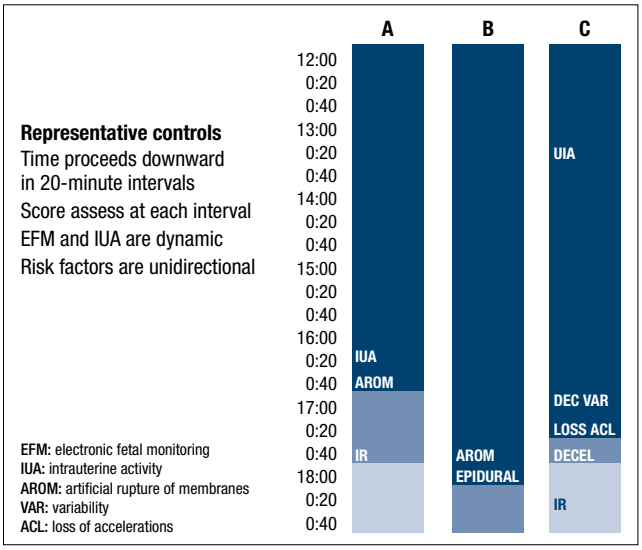
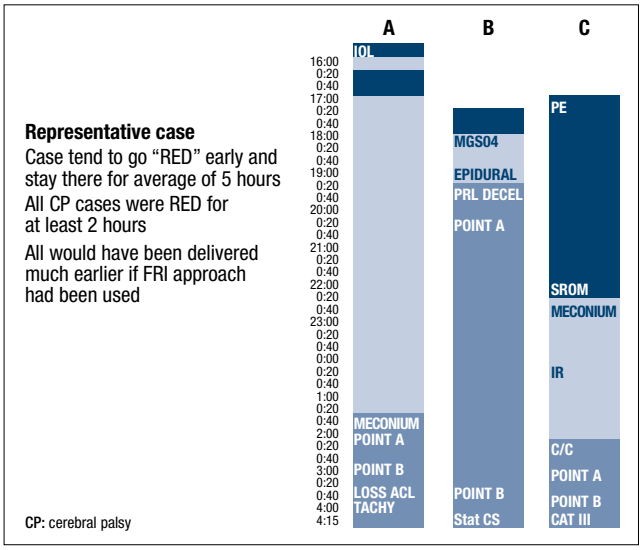


Figure 3 Representative cerebral palsy cases.
Case A: 27 y.o. old primigravida at 41 weeks with hypertension and oligohydramnios admitted for induction of labor. Onset of yellow zone after 11 hours with IUA, decreased variability. Entered red zone with meconium, late recovery with decelerations, onset of 2nd stage pushing, then lost FHR accelerations and developed tachycardia before delivery. No IR performed; never reached CAT III. Apgar's 4/7, pH 6.99, BE -17.
Case B: 21 y.o. primigravida at 40 weeks with obesity, PIH, and meconium. Turned yellow with start of MgS04 with decreased variability, absent FHR accelerations. Red zone entered with epidural, onset of prolonged decelerations, and IUA but no IR performed. Five hours later stat c/section performed; birth weight 3220 g, Apgar's 1/5/7, pH 7.18, BE -14.4.
Case C: a 21 y.o. primigravida at 36 weeks with pre-eclampsia oligohydramnios, SROM and meconium. Red zone entered with onset of late decelerations, loss of variability and accelerations without IR being performed. Terminal deceleration noted. NSVD; birth weight 3050 g, Apgar's 0/0/3, pH 7.13, BE 6.00. Developed hypoxic ischemic encephalopathy (HIE) and seizures within 12 hours of delivery.



As stated previously, all CP cases were "red" for at least 2 hours unless a sudden sentinel event occurred (i.e. prolapsed cord, sudden bradycardia) in which case, the "shot clock" protocol would have ensured patients were delivered well before the 2-hour threshold for CP damage, as we have seen in our previous studies.

Head-to-head analysis of the same cases showed that the sensitivity obtained using the ACOG criteria (pretending we knew prenatally what could only be determined postnatally) was 28%, CAT III had a sensitivity value of 45%, while for the FRI it was 100%. While ithe FRI will never stay at 100%, it was substantially better than existing methods [21]. We now have six published studies with over 1500 control patients and continue to show that the FRI has far better performance metrics than the CAT system [20-24, 28]. Overall, performing a meta-analysis of our publications and combined database, the FRI strongly outperformed CAT III (Table 5) [20-24, 28].

We have also been able to study other aspects of care. While the prediction and prevention of fetal neurological injury are of utmost importance, the incidence of emergency deliveries (EODs) is much higher and takes its own toll on patients, families, and the entire labor and delivery staff [23]. It is well appreciated that such emergency interventions have higher complication rates and exact a "price" even when everything turns out well. Our data show the FRI can anticipate the need for EODs in that, compared with controls, those needing EOD spent an average of 1 hour in the red zone. Among the cases that did not need EOD, most never reached the red zone or were there for a much shorter period of time.

We also performed an intervention series. For 400 cases, management was conducted as per usual clinical routine. Then, one of us applied the principles of the FRI to management and found that the rate of emergency deliveries was reduced from 17% to 4% (65%), emergency cesarean deliveries decreased from 8.5% to 3.3% (62%) as the utilization of IR more than doubled (20% to 47%) [22].

These findings suggest that one of the principal benefits of the FRI is earlier identification of problems that have a higher likelihood of being neutralized by earlier attention. In another study, we demonstrated that the interaction effect of understanding the risk as the FRI is a capable predictor of the chances of EODs because it accurately identifies the level of malleable risk [24]. When used as a clinical tool, the FRI increases the chances of using IR in stage 1, thereby matching risk and resources, which dramatically reduces the chances of EOD. On the other hand, the earlier one responds to risk signals calling for IR, the better, but this applies only if the risk is high [24].

Table 5 Prediction of cerebral palsy.

PREDICTION OF CEREBRAL PALSY		
	FRI	CAT III
Sensitivity	100	44
Specificity	78	78
PPV	53	33
NPV	100	85
PLR	4.55	2.00
NLR	0.00	0.72

Implications and expectations

Our studies suggest the FRI provides a more reliable metric for assessing risks of fetal compromise and the need for emergency intervention than those currently provided by existing methods of EFM interpretation. The CAT system is much too complex and subjective for front line management. There are too many variables that have to be informally considered, and there is no clear, straightforward method of management. Anecdotally, some experienced fetal medicine specialists have responded to our system stating that they do not need the FRI because they have always factored in "other factors" in their interpretation. Unfortunately, many physicians are not sufficiently capable of such expert subjective judgements, necessary to overcome the limitations of the CAT system. A good analogy is the diagnosis of myocardial infarction. For decades, the diagnosis was a gestalt incorporating clinical signs and symptoms, interpretation of the ECG, and non-specific blood tests. It was the discovery of the CPK isoenzymes in the 1970s (and later troponin) that turned the diagnosis into a lab test that had considerably improved metrics [6].

In developing the FRI, attention has been paid to each of these issues, the most important of which, we believe, is the notion of the role of EFM in avoiding fetal harm, which includes the need to avoid an emergency delivery during a trial of labor [23]. We have attempted to change the objective of surveillance from trying to decide the severity of asphyxia and "rescue" to "keeping the fetus out of harm's way in the first place". We do this by changing the mindset, switching from focusing attention on diagnosis of the severity of acidemia to instead recognizing the state at which the fetus cannot be guaranteed to be normal but is not yet definitively damaged. As such, EFM parameters can be used as screening criteria before fetal neurological injury actually occurs.

There is a typical pattern of FHR changes and FRI scores as the clinical situation in labor worsens. The parameters (heart rate, variability, accelerations, and decelerations) do not change independently of one another, and the order of EFM deterioration and occurrence of labor events, (e.g. meconium, 2nd stage, need for IR) is not random. Anticipating pathophysiological deterioration, the red zone is reached when at least two of the EFM screening test variables are still normal. This earlier "warning alarm," i.e. the identification of problems earlier in the pathophysiology, is the critical difference between the FRI and the CAT system, as the former generally allows more time for IR to attempt to halt the progression.

We treat the EFM tracing as a language, albeit an imperfect one to be sure. We use this "language" to query the fetus, not asking, "What is your pH? but, rather "How did you like that contraction?" This approach begins at the onset of monitoring. We use the observed pattern to define whether the FRI can distinguish between cases deemed normal vs abnormal on admission. Our approaches in our published studies have, to date, focused on how behaviorally normal neurologically intact

fetuses respond to the stresses and events of labor. If the fetus is determined upon admission to be already compromised, then a different set of management approaches apply (not discussed here).

Our studies to date have progressed substantially through the phase of development towards establishing proof of principle. Automation is underway to make our approach practical for frontline use. Then will come large scale studies using data in an electronic medical records format. Finally, there will be live implementation.

A priori, we attempt to safeguard the fetus by paying attention to a lowered FRI due to IUA, especially in the 2nd stage. This is independent of FHR patterns. Similarly, the FRI is lowered when data are missing. We believe it is safer to assume "abnormal" and have data to refute, rather than vice versa. We have previously shown that by focusing on limited interventions earlier in the course of fetal deterioration (especially those involved with pushing in the 2nd stage of labor), we can diminish the need to rescue the fetus for heart rate patterns when significant fetal injury has occurred before labor, and no fetal hypoxia or acidosis is currently present.

Conceptually, our philosophy is that EFM is only a screening test, not a diagnostic one. It must be treated as if it were a lab test with a single score. The more subjectivity there is in reaching a conclusion, the less precise any screening classification will be. The military has its weapon systems designed by geniuses, but they have to be capable of being operated successfully by high school only educated troops. EFM as practiced has failed miserably in this sense. It is time to recognize that just as highly experienced commercial airline pilots routinely use computer directed/assisted landing programs, even experienced obstetricians can benefit from computer assisted management of the complexities of labor and heart rate patterns.

Generalizable concepts from our approach suggest the need to see the "big picture" first and then cone down to specific circumstances. Over a decade ago, we suggested that the protocols of tertiary referrals were generally backwards, i.e. a less trained provider is deciding if a patient should be triaged upwards [29, 30]. We continue to believe and have shown that higher level evaluation and appropriate triage downward produces better, and likely cheaper, care.

This is consistent with the "inverted pyramid" that Nicolaides et al. later suggested for prenatal care [31] The realities of medical care in the current environment require better and cheaper approaches. We must continue to develop technologies to help providers make more accurate assessments of risks to empower earlier interventions.

References

1. Cohen A, Hanft R. Policy Directions for effctive evaluation and management. In: Hanft ABCaRS, editor. Technology in American Health Care. Ann Arbor, MI: University of Michigan Press; 2004: 480.
2. Evans M, Hanft R. The Introduction of new technologies. ACOG Clinical Seminars. Washington: ACOG;1997:3.
3. Kuhn T. The Structure of Scientific Revolutions. 4th ed. Chicago: University of Chicago Press; 1992:267.
4. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. N Engl J Med. 1971;284:878-81.
5. Leveton LB, Sox HC, Jr., Stoto MA. HIV and the blood supply: an analysis of crisis decision making. Executive summary. The Institute of Medicine, National Academy of Sciences Committee to Study HIV Transmission Through Blood and Blood Products. Transfusion. 1996;36:919-27.
6. Galen RS, Reiffel JA, Gambino R. Diagnosis of acute myocardial infarction. Relative efficiency of serum enzyme and isoenzyme measurements. JAMA. 1975;232:145-7.
7. Evans MI, Evans SM, Bennett TA, Wapner RJ. The price of abandoning diagnostic testing for cell-free fetal DNA screening. Prenat Diagn. 2018;38:243-5.
8. Evans MI, Andriole S, Curtis J, Evans SM, Kessler AA, Rubenstein AF. The epidemic of abnormal copy number variant cases missed because of reliance upon noninvasive prenatal screening. Prenat Diagn. 2018;38:730-4.
9. Haque IS, Lazarin GA, Kang HP, Evans EA, Goldberg JD, Wapner RJ. Modeled Fetal Risk of Genetic Diseases Identified by Expanded Carrier Screening. JAMA. 2016;316:734-42.
10. Evans MI, Vermeesch JR. Current controversies in prenatal diagnosis 3: industry drives innovation in research and clinical application of genetic prenatal diagnosis and screening. Prenat Diagn. 2016;36:1172-7.
11. Grimes DA, Peipert JF. Electronic fetal monitoring as a public health screening program: the arithmetic of failure. Obstet Gynecol. 2010; 116:1397-400.
12. Evans M, Krivchenia, X, Wapner R, Depp R. Principles of Screening. In: (ed) EM, editor. Clinical Obstetrics & Gynecology. Philadelphia: Lippincott, Williams & Wilkins; 2002:657-60.
13. Galen R, Gambino R. Beyond Normality: Wiley biomedical publication; 1975:237.
14. Macones GA, Hankins GD, Spong CY, Hauth J, Moore T. The 2008 National Institute of Child Health and Human Development workshop report on electronic fetal monitoring: update on definitions, interpretation, and research guidelines. J Obstet Gynecol Neonatal Nurs. 2008;37:510-5.
15. Clark E. Cardiotocography alone is outdated and ST analysis is the way forward in fetal monitoring: AGAINST: An opportunity to avoid past mistakes in fetal monitoring. BJOG 2016;123:1637.
16. Clark SL, Hamilton E, Garite TJ, Timmins A, Warrick PA, Smith S. The Limits of Electronic Fetal Heart Rate Monitoring in the Prevention of Neonatal Metabolic Acidemia. Am J Obstet Gynecol. 2017;216:163.e1-163.e6.
17. Timmins AE, Clark SL. How to Approach Intrapartum Category II Tracings. Obstet Gynecol Clin North Am. 2015;42:363-75.
18. ACOG. Neonatal Encephalopathy and Neurologic Outcome. D'Alton M, editor: ACOG; 2014.
19. ACOG. Neonatal Encephalopathy and Cerebral Palsy: Defining the pathogenesis and Pathophysiology. Washington, DC: ACOG; January, 2003:94.
20. Evans MI, Eden RD, Britt DW, Evans SM, Schifrin BS. Re-engineering the interpretation of electronic fetal monitoring to identify reversible risk for cerebral palsy: a case control series. J Matern Fetal Neonatal Med. 2019;32:2561-9.
21. Eden RD, Evans MI, Evans SM, Schifrin BS. The "fetal reserve index": re-engineering the interpretation and responses to fetal heart rate patterns. Fetal Diagn Ther. 2018;43:90-104.
22. Eden RD, Evans MI, Britt DW, Evans SM, Schifrin BS. Safely lowering the emergency Cesarean and operative vaginal delivery rates using the Fetal Reserve Index. J Matern Fetal Neonatal Med. 2018:1-7.
23. Eden RD, Evans MI, Evans SM, Schifrin BS. Reengineering Electronic Fetal Monitoring Interpretation: Using the Fetal Reserve Index to Anticipate the Need for Emergent Operative Delivery. Reprod Sci. 2018;25:487-97.
24. Britt DW, Evans MI, Schifrin BS, Eden RD. Refining the prediction and prevention of emergency operative deliveries with the fetal reserve index. Fetal Diagn Ther. 2018:1-7.

25. Freeman RK, Goebelsman UW, Nochimson D, Cetrulo C. An evaluation of the significance of positive oxytocin challenge test. Obstet Gynecol. 1976;47:8-13.
26. Schifrin BS, Artenos J, Lyseight N. Late-onset fetal cardiac decelerations associated with fetal breathing movements. J Matern Fetal Neonatal Med. 2002;12:253-9.
27. Evans M, O'Brien J, Dvorin E, Moody J, Drugan A. Alpha fetoprotein. In: (ed) EM, editor. Reproductive Risk and Prenatal Diagnosis. Appleton & Lange Publ Co, Norwalk, CT: 1992.
28. Evans M, Britt D, Eden R, Gallagher P, Evans S, Schifrin BS. The Fetal Reserve Index Significantly Outperforms ACOG Category System in Predicting Cord Blood Base Excess and pH: a methodological failure of the Category System. Reprod Sci. 2019:1933719119833796.
29. Eden RD, Penka A, Britt DW, Landsberger EJ, Evans MI. Re-evaluating the role of the MFM specialist: lead, follow, or get out of the way. J Matern Fetal Neonatal Med. 2005;18:253-8.
30. Britt DW, Eden RD, Evans MI. Matching risk and resources in high-risk pregnancies. J Matern Fetal Neonatal Med. 2006;19:645-50.
31. Nicolaides KH. Turning the pyramid of prenatal care. Fetal Diagn Ther. 2011;29:183-96.

Conflicts of interest: *The author declares that there is no conflict of interest*