



# Changing Perspectives of Electronic Fetal Monitoring

Mark I. Evans<sup>1,2,3</sup> · David W. Britt<sup>1</sup> · Shara M. Evans<sup>4</sup> · Lawrence D. Devoe<sup>5</sup>

Received: 19 July 2021 / Accepted: 24 September 2021  
© Society for Reproductive Investigation 2021

## Abstract

The delivery of healthy babies is the primary goal of obstetric care. Many technologies have been developed to reduce both maternal and fetal risks for poor outcomes. For 50 years, electronic fetal monitoring (EFM) has been used extensively in labor attempting to prevent a large proportion of neonatal encephalopathy and cerebral palsy. However, even key opinion leaders admit that EFM has mostly failed to achieve this goal. We believe this situation emanates from a fundamental misunderstanding of differences between screening and diagnostic tests, considerable subjectivity and inter-observer variability in EFM interpretation, failure to address the pathophysiology of fetal compromise, and a tunnel vision focus. To address these suboptimal results, several iterations of increasingly sophisticated analyses have intended to improve the situation. We believe that part of the continuing problem is that the focus of EFM has been too narrow ignoring important contextual issues such as maternal, fetal, and obstetrical risk factors, and increased uterine contraction frequency. All of these can significantly impact the application of EFM to intrapartum care. We have recently developed a new clinical approach, the Fetal Reserve Index (FRI), contextualizing EFM interpretation. Our data suggest the FRI is capable of providing higher accuracy and earlier detection of emerging fetal compromise. Over time, artificial intelligence/machine learning approaches will likely improve measurements and interpretation of FHR characteristics and other relevant variables. Such future developments will allow us to develop more comprehensive models that increase the interpretability and utility of interfaces for clinical decision making during the intrapartum period.

**Keywords** Electronic fetal monitoring · Screening tests · Artificial intelligence · Computerized electronic fetal monitoring · Fetal Reserve Index

## Fundamentals

In 1889, there was a serious proposal to close down the United States Patent Office because “everything” that can be invented had already been invented [1]. Roughly 11 million issued patents later, the submission rate continues to be

robust, and filings based upon foreign issuances and submissions form an ever-increasing proportion of the Patent Office’s work. Clearly, the process of creating new knowledge, publishing it, and protecting it continues to progress at an ever-increasing rate. The sources of support for such work have expanded over the past couple of decades as traditional funders of such efforts (universities, governments, and philanthropists) have been supplemented and, in some cases, replaced by venture capitalists whose business models are not known for their patience.

In many respects, the old adage “necessity is the mother of invention” has been replaced by “invention is the mother of necessity” [2]. The Fax machine in the 1980s, cell phones and email in the 90 s, I-pads in the 00 s, and the mushrooming of Zoom and related technologies in the pandemic are classic examples of technological capabilities creating a need for their use. This is easy to overstate, however. For example, Motorola developed a “game changing” new satellite phone in the 1990s. However, it had a fundamental flaw.

---

✉ Mark I. Evans  
evans@compregen.com

<sup>1</sup> Fetal Medicine Foundation of America, New York, NY, USA

<sup>2</sup> Comprehensive Genetics, PLLC, New York, NY, USA

<sup>3</sup> Department of Obstetrics & Gynecology, Icahn School of Medicine at Mt. Sinai, New York, NY, USA

<sup>4</sup> Department of Maternal Child Health, Gillings School of Public Health, University of North Carolina, Chapel Hill, USA

<sup>5</sup> Department of Obstetrics and Gynecology, Medical College of Georgia, Augusta University, Augusta, GA, USA

It weighed as much as a brick and cost a small fortune. One of the lessons that came out of an NSF analysis of adoption of new communication technologies was that they had to be both reliable and have real, perceived, and actual advantages for widescale adoption to take place [3].

In medicine, there is inherent stress emanating from pressures to produce new technologies that improve care and, at the same time, are novel enough to create a niche in the marketplace. However, historically, physicians, and particularly obstetricians, have been far more hesitant to adopt “game changing,” disruptive technologies than are professionals working in many other consumer products, services, and scientific fields [2]. Here, we will review the 50-year history of electronic fetal monitoring (EFM) in the USA, known as cardiotocography (CTG) in much of the rest of the world. We believe that, for purposes of evaluation, EFM must be construed as a classic screening test. Only then can we assess how actual technology approaches both followed and skirted basic evaluation principles for development and diffusion, how new approaches were implemented, how methods used and ignored classic statistical performance metrics, and how attention has been fixated on FHR characteristics rather than developing more comprehensive, contextualized approaches [4].

There are many reasons for the relatively slow adoption rates of new technologies in medicine. With every potential for significant improvements in patient care and their attendant rewards (academic, personal, and sometimes financial) comes a negative potential risk for being wrong (e.g., drugs such as diethylstilbestrol and technologies such as fetal oximetry) [5, 6]. In the USA and several other countries, fear of liability exposure, which is very high in obstetrics as compared to many other medical fields, is often cited as the driving force that limits care options [7, 8]. It often appears that only after a new approach has become widely accepted throughout much of the world does it enter into practice in the USA.

There is a dedicated field of technology innovation and assessment whose principles are generalizable across disciplines. This field has its own literature, societies, and meetings that, in turn, create norms, models, and expectations [9, 10]. An understanding of the commonalities of introducing new technology, regardless of the specific field, can help to avoid common pitfalls in such undertakings.

We can divide technology development broadly into two distinct components. First there is the phase of “development.” In general, and more commonly in the past, new ideas in medicine most often originated within academic settings, where they were conceptualized, tested, published, and occasionally patented [9, 10]. Such publications and presentations often begin to create demand for the application of these new ideas. As such demand increases, the original developers can no longer manage it. Other parties

want “in” on the game, and new users enter the field. This is the “diffusion” phase. As classic technology assessment literature has shown for decades, during diffusion, utilization of new technologies increases, and complications skyrocket. Only after a period of time during which clinicians learn the new technology, often led by younger practitioners who did not “grow up” using the old model, does the situation calm down. These same concepts apply equally to medical therapies, surgical procedures, and laboratory tests [9, 10].

The life cycle of a new or better technology eventually results in its displacing the old technology. This situation is often complicated by the behavior of those accustomed to the existing technology who may be motivated in terms of beliefs (claims to superior knowledge) or by financial interests (sources of their livelihood) to preserve the existing technology and undermine emerging rival technologies. The Nobel Laureate, Max Planck, presented a dismal view of this process more than a hundred years ago, arguing that “a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it” [11].

Screening for risk and testing for actual diagnoses are integral parts of medical practice. When compared with many other medical specialties, Obstetrics usually has a higher percentage of patients who are healthy. This specialty tends to utilize screening tests more often than other specialties to cull the larger mostly low-risk population. The goal is to identify a subset of high-risk patients for whom the use of more expensive diagnostic tests is justified that may then lead to riskier interventions and subsequent complications. Many screening and diagnostic tests have often adhered to the fundamental criteria (below), but some have deviated considerably from them and have yielded highly variable results.

There are several criteria that are generally felt required before deciding to screen for a condition (Table 1) [12]. The principles of evaluation, or performance characteristics, were introduced into clinical practice in the 1970s by Galen and Gambino [13]. They establish the boundaries of a

**Table 1** Criteria for program screening

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

playing field and a scoring system within which competitors that may offer better ways of doing things can be evaluated. Lack of rigor in development and application of screening tests can lead to poor performance metrics (e.g., low sensitivity and positive predictive values). Disproportionate reliance, expectations, expenditures, and complications can result from follow-up diagnostic testing that was likely unnecessary. An often-underappreciated goal of a screening program that is its ability to detect the maximum number of affected individuals for the lowest number that screen positive in a population, rather than to identify any specific individual with the condition or problem of interest [13]. Where to put the cutoff point is arbitrary, but it must be maintained to ensure maximum precision.

Many, if not most, patients and physicians do not understand the difference between screening and diagnosis [4, 13]. Screening tests are meant for all patients, and they are only required to distinguish individuals with high enough risk to warrant diagnostic testing from those without such risk. They do not give definitive answers. How well they do their job is defined by the metrics of sensitivity, specificity, and positive and negative predictive values [4, 12]. Diagnostic tests are meant to give definitive answers. Such tests may carry risks, may be expensive, and consequently are only intended for patients at a high enough risk to warrant their use.

In Obstetrics, many of the outcomes of concern are rare, so looking only at the overall accuracy of a test or a model constructed from logistic regression (the proportion of all cases that are correctly identified as being either problematic or unproblematic) is almost useless. This is because it will identify so many true negatives because of the rarity of the phenomenon in question. Sensitivity, known more generally as “Recall” for those working in Machine Learning, has become standard when the cost of being wrong is high. The primary question of interest is: of all the cases that are considered problematic by the screen, what proportion are actually problematic? Positive Predictive Value (PPV), also known as “Precision,” helps us understand how correct a model is when it predicts there is a problem: it is the proportion of all cases predicted as problematic that are actually problematic [14, 15]. Finally, a summary score of Accuracy that is replacing the traditional measure of accuracy is called the F1 score. It combines both Precision and Recall in such a way as to yield a metric that varies between 0 and 1, with low scores representing a test that yields large proportions of false positives and false negatives, and high scores (closer to 1) identifying tests or models that have very low percentages of false positives and false negatives.

The above principles lay the foundation to analyze how the impact of technology like EFM could best be evaluated. By analogy, currently debated evolutions include the use of cell-free fetal DNA (cffDNA) versus procedures

with microarrays, and pan-ethnic Mendelian screening [1, 16–18]; cffDNA clearly identifies an increased percentage of fetuses with Down syndrome, but at the cost of abandonment of diagnostic procedures with microarray analysis which could detect a far higher number of additional serious disorders [14, 15]. Conversely, pan-ethnic Mendelian screening is still underutilized [16]. 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 [16].

We believe a considerable reason for the disparities between utilization of cffDNA versus utilization of microarrays and expanded Mendelian screening is related to the culture of venture capitalism vs traditional academic underpinnings [1]. cffDNA was rushed into practice with high marketing budget pressures by companies but relatively few peer-reviewed publications. Microarrays followed the traditional rigorous process of undergoing numerous studies, including an NIH-sponsored multicenter randomized trial, before being entered into practice [16, 17, 19]. Similarly, the rapid adoption of EFM into clinical practice occurred before many basic principles for its application had been established and properly understood [20].

## Moving Goal Posts

We have previously written on many of the basic principles of screening and argued that EFM as it is currently being used falls short of satisfying them [4, 21–30]. In spite of being introduced into widespread practice about 50 years ago, there is still a wide confusion as to precisely what EFM is trying to accomplish, and how well it does it [20].

We have already presented what a screening test is intended to accomplish and that there should be an a priori consensus on the definitions of what is normal and what is abnormal. If an a priori consensus has not been established, then the ability of any screening test to distinguish between these two conditions will be impossible [4].

Today, nearly all obstetric care providers can agree that the use of EFM has succeeded in significantly reducing the risk of preventable intrapartum fetal deaths [31, 32]. If that were the only goal and metric, there would be nothing further to discuss. However, a subsequent goal for EFM was to try to prevent fetal hypoxia, fetal acidemia, neonatal encephalopathy, and cerebral palsy (CP). If one considers CP as the final “dependent variable,” all of the other variables, while being diagnoses in their own right, can also be thought of as screening tests for CP or least as alternative pathways of a model for health assessment. In the 1970s and early 1980s, when fetal scalp blood sampling (FBS) was common, acidemia could be directly evaluated by measuring capillary blood pH and base excess [33].

Eventually, EFM was felt to be accurate enough to replace FBS (not entirely true in Europe where EFM developed in parallel with the USA and South America). Given the generally accepted conclusion that EFM does not perform well enough, it is not surprising that some investigators are still trying to improve it [34, 35].

At its inception, EFM interpretation was a cottage industry with wide variation in what was considered normal and abnormal and a wide variety of terms and definitions for interpretation. The NICHD-initiated effort in the US to improve this confusing situation was to organize an expert panel that would create a uniform terminology for EFM interpretation [34]. This was followed up by a second expert panel that generated an ACOG Practice Bulletin to present a summary of the uniform terminology for EFM interpretation [36].

In response to the challenges of preventing the adverse outcomes of neonatal encephalopathy and CP, in 2000, the American College of Obstetricians and Gynecologists (ACOG) assembled an interdisciplinary committee. This committee's charge was to review the literature and establish a compendium of the best science available to determine the relative associations of intrapartum factors with the outcomes of neonatal encephalopathy and CP. A monograph was published in 2003 which concluded that most CP was preordained before labor—either from genetic or other preexisting conditions. As a result, most CP cases were thought not to be caused by events during labor and delivery and were essentially nonpreventable [31] (Table 2).

This first monograph helped to develop a way of thinking about causation in the development of CP. It did not solve all the problems surrounding the onset or causation of this condition and provided little help for directing management approaches. A second edition in 2014 updated

the first edition but reached very similar conclusions (discussed below) [32].

## How Much of Cerebral Palsy is Preordained?

A very high proportion of obstetrical malpractice litigation has focused on whether an allegedly damaged baby developed CP because of mismanagement of labor and delivery or whether it was caused by events or conditions that preceded the onset of labor or came after birth [37]. A complete discussion of this issue is beyond the scope of this paper, but some salient concepts have been developed over the past several years. The extremist notions that either all cases of CP are genetic or that none are, have fortunately become less frequently promoted but have not entirely disappeared. Even the conceptual shift that most CP occurred before labor began must be further subdivided into those causes, such as genetic, that are not preventable, and those that might have been prevented with earlier recognition of risk [38]. The simplistic idea, in part promulgated by the ACOG monographs, was that rigid criteria (e.g., a pH in cord blood < 7.00) were required to attribute CP to intrapartum events. This notion has been refuted by more recent data that have found that most CP cases never exhibited Category III tracings which were thought to reflect such conditions [21]. Nevertheless, such polarized opinions are still baked into much of the discussion (discussed below).

## Genetics of Cerebral Palsy

For decades, reports, such as that by Nelson and Ellenberg in 1984, studied the incidence of CP in term babies with and without obstetrical complications and with and without febrility [39]. They reported essentially no difference in the incidence of CP in babies > 2500 g regardless of the presence of obstetrical complications as long as the 5-min

**Table 2** ACOG essential criteria

ACOG essential criteria to conclude NE related to “an acute intrapartum event” (must meet all four)

1. Evidence of a metabolic acidosis in fetal umbilical cord arterial blood obtained at delivery (pH < 7 and base deficit  $\geq$  12 mMol/L)
2. Early onset of severe or moderate neonatal encephalopathy in infants born at 34 or more weeks of gestation
3. Cerebral palsy of the spastic quadriplegic or dyskinetic type
4. Exclusion of other identifiable etiologies such as trauma, coagulation disorders, infectious conditions, or “genetic disorders”

Criteria that collectively suggest an intrapartum timing (within close proximity to labor and delivery (e.g., 0–48 h) but are nonspecific as to aphyxial insults

1. **A sentinel (signal) hypoxic event occurring immediately before or during labor**
2. **A sudden and sustained fetal bradycardia or the absence of fetal heart rate variability in the presence of persistent, late, or variable decelerations, usually after a hypoxic sentinel event when the pattern was previously normal**
3. **Apgar scores of 0–3 beyond 5 min**
4. **Onset of multisystem involvement within 72 h of birth**
5. **Early imaging study showing evidence of acute non-focal “cerebral abnormality”**

Apgar returned to normal. In 1986, they reported that only about 21% of CP children had a marker for asphyxia, and of these more than half had a congenital malformation or other intrinsic defect that might have contributed to an unfavorable outcome [40]. Later, infection—either pre or postnatal—was added to the non-labor-related causative etiologies. Overall, these papers led to the perception (commonly quoted) that perhaps 90% of CP was either of genetic or had some other congenital basis [41].

Numerous genetic etiologies for CP were detailed in the first edition of the ACOG monograph, but those (mostly due to biochemical abnormalities) represented a small proportion of affected cases [29]. The development of increasingly sophisticated laboratory techniques such as microarrays and whole exome sequencing (WES) has enabled more clarification of the molecular etiology of hundreds of types of abnormalities for CP and many other issues [42, 43]. Nakao et al. recently categorized over 1000 cases of CP in Japan [44]. They divided patients into 5 categories: (1) Cases with bradycardia on admission (suggesting damage already occurred), (2) Persistently non-reassuring FHR, (3) Reassuring prolonged decelerations, (4) “Hon’s pattern” of reassuring FHR gradually deteriorating, and (5) Persistently reassuring patterns throughout [44]. Our analysis of their data suggested that 30% cases were determined before admission, which we believe include cases that are due to genetic causes and those that are not but that had suffered damage in the antenatal period [38]. Nakao et al. also found 19% of CP cases were unclassifiable (which by itself is enough to conclude that EFM is not a well-performing screening test). Only 16% of CP babies exhibited “Hon’s pattern” which they concluded were the only ones for which prenatal management might have contributed to neonatal compromise [44]. Recent papers using whole-exome sequencing (WES) have suggested that approximately one third of CP cases have a molecular variant responsible for their phenotypic manifestations [45]. This seems to be a very reasonable conclusion. The exact penetrance and expressivity of these markers will clearly vary, and their relationship with external stresses will need considerable exploration. Furthermore, just as ultrasound markers such as a 2-vessel umbilical cord and choroid plexus cysts were originally felt to carry a very high risk of aneuploidy, improvements in ultrasound examination (and in molecular genetic techniques) now show that those ultrasound markers, when isolated, carry no increased risk of genetic abnormalities [46]. As more CP cases undergo sophisticated genetic testing, it seems quite likely that such cases will sort themselves out—analogue to the advanced investigation of variants of uncertain origin resulting from the new techniques of molecular testing.

Many current papers, such as those by Rossi and Prefumo, started out their investigation of CP and EFM with statements such as “As only a small proportion of HIE

depends upon an asphyxia event during labor, strict criteria were developed, according to which a sentinel event leading to acute fetal hypoxia or placental hypoperfusion is mandatory to classify HIE as a result of intrapartum asphyxia” [31, 32, 47] (Table 2). These criteria are listed in the original ACOG Monograph (2003) [31] of which one of us (MIE) was an author. ACOG published a revised edition in 2014.

Rossi and Prefumo then backed off a bit, noting that “as brain damage might occur in low-risk pregnancies without sentinel events, preconceptual and antepartum risk factors might superimpose to intrapartum hypoxia and predispose the fetus to brain injury secondary to the hypoxia resulting from uterine contractions [48]. There has been a lack of consensus on the exact pathophysiology of damage which we cannot solve here, but it is clear that multiple mechanisms of pathology can only hinder the statistical performance of any single screening and diagnostic approach. Although a healthy fetus is able to sustain the stresses of labor and prolonged second stage, a compromised fetus is highly susceptible to hypoxia generated by uterine contractions. As the timing of hypoxia is hard to ascertain, several studies have attempted to identify risk factors in order to select fetuses that might benefit from strict monitoring during labor [49–51] (discussed later).

In 2008, to refine the intrapartum management aspect of fetal monitoring, ACOG introduced a three-tiered “category system” (CAT system) based on the presumed presence or absence of fetal acidemia [52–54]. 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,” (i.e., non-diagnostic). There is no specific metric or conceptual agreement on how hypoxia or acidosis came to be present, or how much time the fetus has left before irreversible neurological injury occurs. In addition to continued observation, the 2014 ACOG Monograph offers management recommendations for intrapartum resuscitation depending on whether the elements of concern are recurrent variable or late decelerations [32, 54]. Implicit is the assumption that, without acidemia that is significant enough 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 led to improved outcomes [32, 55, 56]. 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 [57]. 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. Most babies with CP do not have CAT III tracings; yet it is a common defense argument that no intervention in labor was needed as the tracings were only CAT II [58].

Screening tests are always a trade-off of increasing sensitivity at the price of more false positives or increasing positive predictive value but tolerating more false negatives. By definition, a screening test is not perfect. If it were, it would be a diagnostic test. By analogy to maternal serum alpha fetoprotein screening for neural tube defects that began in the 1970s, reaching Category III is very far out on the right of the distribution curve such that it has a high positive predictive value. However, there is a poor sensitivity because of too many false negatives [4] (Fig. 1). Developing an optimal cutoff relies upon a combination of the statistical metrics just described but then has to consider the contextualized implications (clinical, financial, legal, and public policy) of being wrong.

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 screening test for injury that has already occurred or is in the process of occurring [31, 32]. By the time CAT III is reached, it is often too late to alter the sequelae of fetal injury, even if an emergency operative delivery (EOD) were performed.

The EFM characteristics of CAT III tracings, which coincide with the critical acid–base derangements noted in the ACOG Monograph criteria, are absent FHR variability or

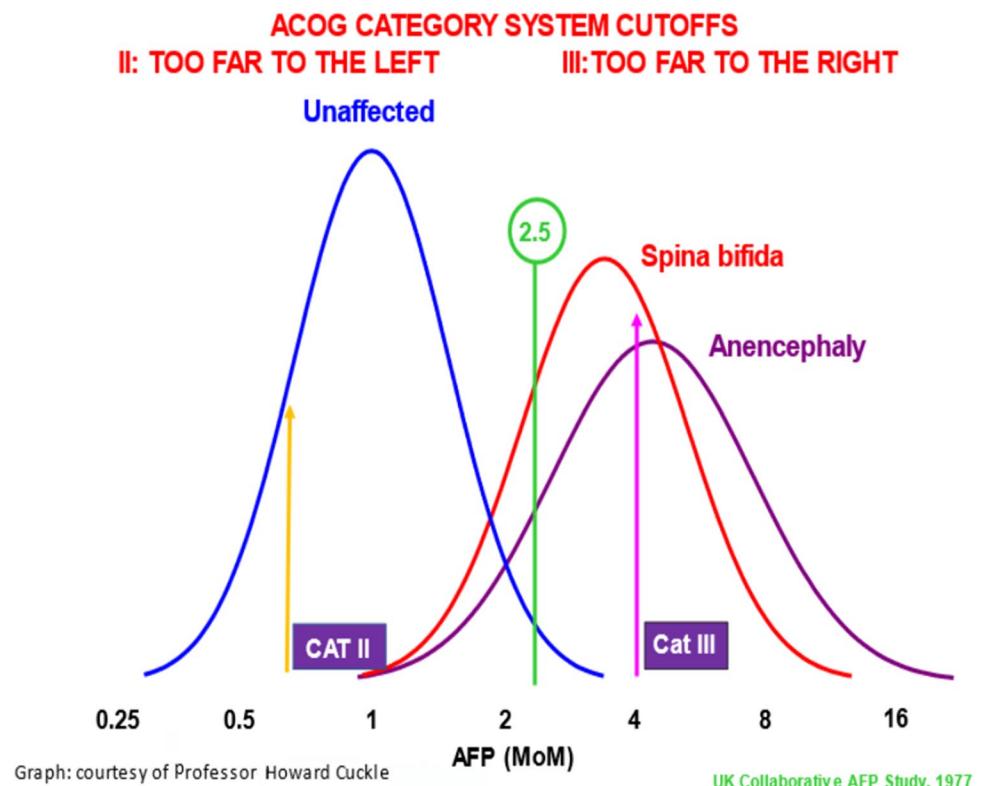
sinusoidal pattern of the FHR baseline, absent FHR accelerations, FHR decelerations with late recovery, absent variability during the recovery and tachycardia (often > 180 bpm), or an agonal baseline.

We believe that in order to prevent neurological injury from occurring, the CAT III diagnostic criteria must be replaced by criteria that would be useful for earlier detection of the sequence of events that leads to abnormal fetal acid–base balance, as well as other mechanisms that may result in neurological injury [4]. Such a screening test could then serve to evaluate the risk of neurological injury if labor were to continue without resuscitation or intervention. Therefore, it is imperative to understand the interrelationships among the EFM monitoring parameters since the presence of one abnormality affects the others and may do so either additively or interactively.

We have analyzed patients who entered labor with CAT I tracings and delivered a baby with CP without any apparent cause other than labor [21–26]. We have 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 that were normal at the onset of labor and that went on to develop CP typically demonstrate a characteristic pattern: hypoxia/ischemia, and predictable deterioration to the point of injury in association with excessive uterine activity ( $\geq 8$  UCs/20 min).

The apparent ontogeny of hypoxia/asphyxia in pregnancies where fetuses are “normal/uninjured” at the onset of

**Fig. 1** Using the screening model of maternal serum alpha fetoprotein as a model, Category III has its cutoff far out to the right on the curve such that it would have a very high positive predictive value but many false negatives. Category II is too far to the left—giving a high sensitivity but too many false positives



labor starts with the occurrence of contractions. In our studies, for control patients (good outcomes), the average length of labor was 11.3 h, while for those patients whose infants developed CP, the average length of labor was 17.7 h. There were several other differences in the average time to initial appearance of EFM abnormalities and the order of deterioration of EFM variables, and there was progressive and relatively predictable loss of reassuring characteristics of EFM parameters [4, 21–30]. 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 happened later than Point B in every case, and most often in the 2nd stage of labor, within 20 min of delivery [21]. Such considerations have been discussed more extensively in our previous publications [4, 21–30]. Points A and B require sophisticated interpretation of EFM and consequently are not included in our proposed methodology that will be described later.

Deterioration of previously reassuring EFM parameters should be used to alert caution, to prompt intrauterine resuscitation (IR), and intervention when necessary, rather than wait until there is a CAT III tracing and irreversible fetal neurological injury. 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 other alteration in the

baseline features in response to those uterine contractions. When variability disappears, the fetus has already spent considerable time and effort compensating for impaired oxygen availability/blood flow. The CAT system requires complete absence of variability before the pattern can be called CAT III. CAT ignores the general ontogeny of these changes which are (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 principle in medicine that one cannot treat a condition until it has been diagnosed. Unless potential changes in fetal tracings that are associated with conditions that could lead to neurologic injury are recognized before that injury occurs, there are no options for amelioration by earlier intervention to prevent that injury. Thus, we believe that to be effective in preventing fetal harm, a screening method must ask how well the fetus can tolerate the effects of each contraction on its oxygen supply and how much “reserve” it has left to withstand the subsequent contractions and to deliver this message in real time.

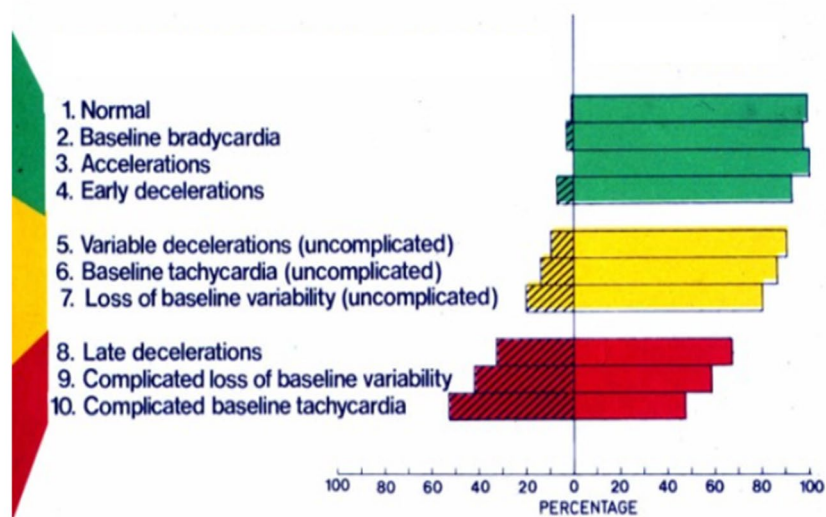
### Automated Systems for EFM Interpretation and Analysis

#### Background

For 50 years, it has generally been accepted that individual abnormal FHR features are predictors of fetal condition

Fig. 2 Association of differing features of electronic fetal monitoring and their impact on predictions of outcomes

## FETAL HEART RATE DETERIORATION AND ACIDOSIS



(see below) (Fig. 2). However, one of the persistent weaknesses of EFM interpretation has been the failure of clinicians to recognize what are and what are not significant abnormalities of FHR patterns. EFM classification schemes from ACOG have attempted to standardize the rating of a number of FHR patterns in prognostic groups ranging from normal (Category I) to intermediate (Category II) to pathological (Category III). These classification schemes are intended to make more uniform and reproducible communication about suspected fetal condition. Despite having been introduced into routine practice over a decade ago, there have been no large-scale studies comparing the 3-tiered (Category system) to the previous 2-tiered classification system to determine if this new system is actually superior to its predecessor for improving fetal outcomes. A Cochrane Review shows that most RCTs have failed to demonstrate that visual FHR interpretation is superior to intermittent auscultation for reducing the rates of perinatal mortality or CP [59]. It should be noted, however, that these RCTs were performed before the development and adoption of the three-tiered CAT system. Of note, such large-scale studies were generally required before major changes in practices were introduced, such as BRCA, preeclampsia screening, chorionic villus sampling, nuchal translucencies, combined prenatal serum and ultrasound screening, and microarrays [60]. Cell free fetal DNA and EFM have been the most prominent exceptions of technologies that gained wide acceptance and utilization before their scientific credibility was established [17, 59].

### Automated Systems for FHR Analysis

Recognizing that visual interpretation is a weak link in EFM's ability to predict fetal condition and that this situation

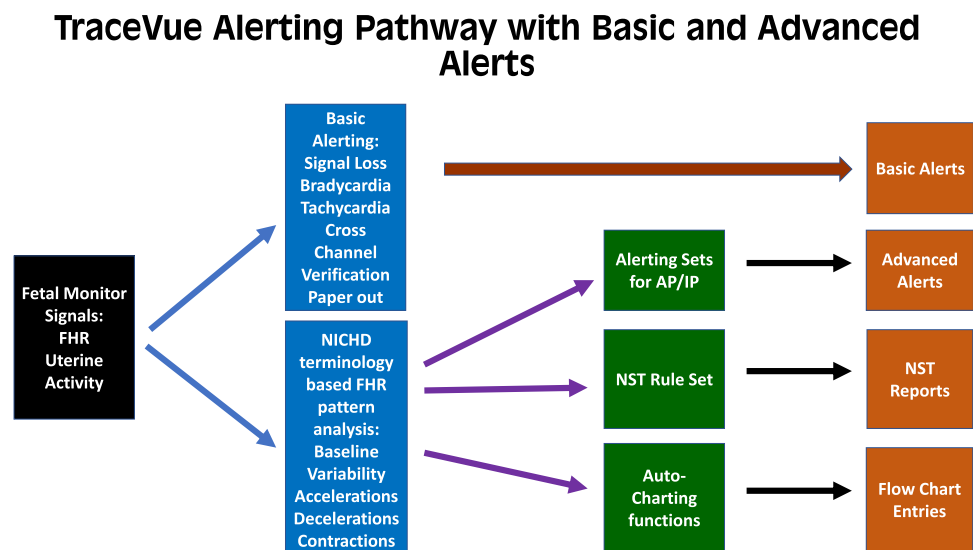
has not changed since its inception, numerous efforts have been made to automate FHR analysis. Historically, these began with Dawes and Redman in Oxford, resulting in the Oxford Sonicaid system [60]. Their system was primarily intended for antenatal assessment to screen for fetal hypoxia or deterioration of fetal status. In this setting, it was successful, but attempts to apply the system to the intrapartum environment consistently failed to discriminate acidotic from non-acidotic fetuses [60].

A few years after the Dawes–Redman system was introduced, one member of our group (LDD) independently developed a rule-based antepartum analytic system (NST-EXPERT) that was similar to the Oxford system [61]. NST-EXPERT was used extensively in our antenatal testing unit but not in the labor/delivery unit. No large studies were performed to demonstrate that it improved perinatal outcomes when compared with visual assessment of the resting FHR tracing [60]. However, following the development of this system, a lengthy collaboration with Hewlett Packard Biomedical was developed. By the early 1990s, this group's efforts resulted in the mature TraceVue analytic system for FHR interpretation. TraceVue is a rule-based system for analyzing FHR patterns and is capable of generating a number of clinical alerts based on abnormal FHR features (Fig. 3).

We performed a side-by-side comparison of this system for identification of FHR pattern features and alerting [61]. Observer–computer agreement (> 80%) was high for most of the standard FHR elements and the respective monitor-generated alerts when they were triggered. An updated version of TraceVue is still in use in many obstetric units throughout the USA and Europe [59]. However, there have been no large-scale studies demonstrating the ability of this system to improve neonatal outcomes.

In the mid-1990's, the Plymouth UK CTG research group, led by clinician Keith Greene and engineer Rob Keith, began

**Fig. 3** TraceVue alerting pathways showing basic and advanced formats





to develop a hybrid rule-based/neural network system for intrapartum CTG analysis. This system eventually matured as K2 technologies and has most recently been marketed as INFANT. INFANT is an advanced analytic/alerting system with four levels of assessment ranging from green to red (see below) [62].

A total of 47,062 women in Ireland and the UK were recruited over three-plus years, to the largest randomized controlled trial of computerized CTG analysis ever undertaken. Providers were trained in the use of INFANT as a decision support system for assisting obstetric care. Control patients were managed by standard visual interpretation alone (INFANT was masked). The study group had the additional INFANT display with the FHR analysis visible to the obstetric providers. Regardless of whether the INFANT data were available or concealed, there were no differences between control and study groups for the primary outcome, which was a composite measure of adverse neonatal outcomes or a secondary outcome, Parent Report of Children's Abilities—Revised (PARCA-R), a developmental assessment at age two [62].

Over the past two decades, Ayres-de-Campos et al. independently constructed an automated FHR analytic system, SIS-PORTO [63]. This system also created a four-tier approach to FHR categorization for alerting providers. A UK multicenter trial enrolled 7700 patients, but again it failed to show significant differences between study and control groups for adverse neonatal outcomes such as metabolic acidosis, HIE, or death [64].

Elliott et al. developed a proprietary automated FHR analysis system PeriCALM® Tracings™ sold in the USA by Perigen [65]. Its analytic software is based on ACOG Classification of FHR Categories, and it was visually validated by experts in FHR interpretation. However, it has not been rigorously tested for improvement of fetal condition. It employs an artificial intelligence (AI) program to aid in the recognition of potentially pathologic EFM patterns. It also provides visual cues to alert caregivers to abnormal trends in FHR pattern and labor abnormalities. No large, randomized trials of this system have been conducted to date to show improvement in neonatal outcomes.

Adjunctive assessment measures such as FBS, fetal pulse oximetry, and fetal ECG ST-segment analysis (STAN) represented efforts to improve the assessment of intrapartum fetal condition when the FHR pattern was being scrutinized. FBS has been abandoned in the USA [65]. Likewise, fetal pulse oximetry was also abandoned when it was shown in a large multi-center study not to reduce cesarean delivery rates or improve neonatal outcomes [66].

Animal studies demonstrated that in the presence of hypoxia, there was considerable elevation of the T-wave of the ECG. As the fetus transitions from aerobic to anaerobic metabolism, the ratio of T-wave height to QRS complex

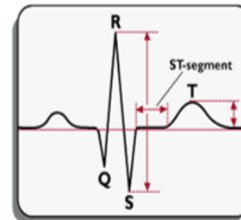
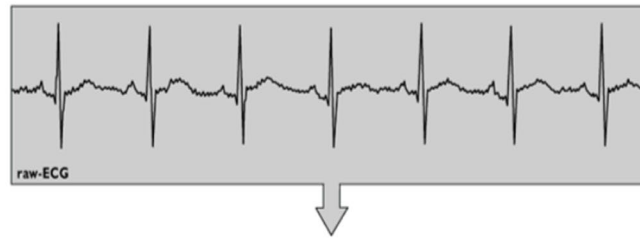
amplitude is elevated, leading to the concept of the T:QRS ratio. As capture of the fetal ECG improved, these data were more reliably detected, and clinical guidelines for applying this metric to intrapartum fetal assessment were developed [67]. The result of this research was Neoventa Medical's STAN monitoring system that coupled a standard fetal monitor with fetal ECG analysis. Coupled with a four-tiered FHR (three-tiered FHR in the USA) classification system, a clinical management scheme that combines the fetal ECG analysis with FHR patterns was developed. The recording of the fetal ECG is noted below, and abnormalities in ST wave and T:QRS ratios were among the so-called ST events that formed the basis for alerts. Depending on the category of the FHR tracing, either continued observation or some form of obstetric intervention would be suggested by the STAN Guidelines. Such interventions could be in the form of intrauterine resuscitation and/or expeditious delivery [67] (Figs. 4, 5, and 6).

There have been several randomized controlled trials of the STAN system beginning with one in the UK and the other in Sweden [68, 69]. Several meta-analyses of these and three more RCT studies have subsequently been performed. One of us (LDD) participated in a reanalysis of the RCTs and their previous meta-analyses. The analysis found enough differences between the conduct of these studies and their reports to show that most were underpowered for the critical primary outcomes such as neonatal metabolic acidosis, HIE, and mortality. In contrast, the majority of properly done STAN trials that adhered to the established clinical protocols did show a reduction in neonatal metabolic acidosis, use of fetal scalp blood sampling, and rates of operative delivery [70, 71].

The final blow to the implementation to the STAN methodology in the USA was delivered by a large RCT conducted by the Maternal–Fetal Medicine Unit Network [72]. A total of 11,108 patients were randomized into two groups. The primary composite outcome of adverse neonatal outcome or secondary outcomes of route of delivery did not differ between the two groups. Important criticisms of this study included the observation that many of the patients enrolled were not as high risk as those enrolled in previous RCTs of the STAN methodology. However, based on this study's findings and conclusions, STAN is no longer being employed in US labor units. A more recent study using ST waveform analysis performed on nearly 20,000 patients at a single center in the Netherlands showed marked diminishment of metabolic acidosis and neonatal encephalopathy [73]. While this was a retrospective longitudinal study, it did show that there was a significant learning curve in adopting this new technology as the outcomes in the second half of the study were notably better than those of its first half. Because there was no true control group, it was not possible

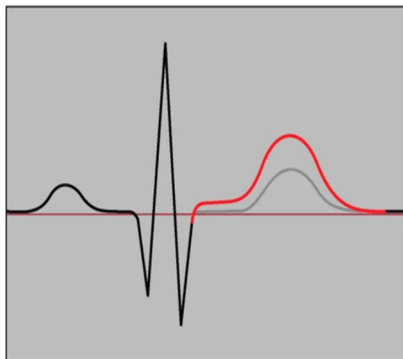
**Fig. 4** Measurement of QRS and T wave amplitude as the basis for T:QRS ratio

## Measurement of QRS amplitude and T-wave height



NeoventaMedical AB 2009

## Effect of Hypoxia on T-wave height



NeoventaMedical AB 2009

**Fig. 5** The effect of hypoxia on T wave amplitude

to attribute these improvements in neonatal outcome solely to the addition of ST analysis.

The conundrum here is that this study should fall under the phase of “development.” Yet, it was actually presented as an implementation one. To do a study such as this properly requires the full commitment of an obstetrical unit’s staff to the required specialized training and regular quality reviews, continuous feedback, continuing reinforcement, and education. Parenthetically, one member of our group (LDD) participated in the first US multi-center clinical trial of the STAN system [74]. We demonstrated successful incorporation into six obstetric units across the USA. There was a high degree of correlation between the intervention and nonintervention

## Category II FHR with ST events



NeoventaMedical AB 2009

**Fig. 6** A Category II tracing with ST event alerts

decisions made by the American clinicians and the three STAN experts who independently conducted retrospective case blinded reviews. No fetus found to have metabolic acidosis was missed. While this study was limited in size (530 high-risk patients were enrolled), it did demonstrate the feasibility of transporting a new technology-based EFM interpretative system from its country of origin (Sweden) to the USA.

We contend that this technology did not successfully diffuse out into the community to improve overall public health because the level of expertise and compulsiveness required to “operate” it were beyond the capabilities of the average practitioner. Thus, its failure was, in part, an implementation issue. Implementation science encompasses the spectrum

of events that needs to occur in order to make the transition from the “Bench to the Bedside.” Like a hurdle race, getting over all of them is required to have a chance to win; any one failure will sink the entire enterprise. The Landsman STAN study demonstrated an improvement in outcomes when it was in superb hands. However, what was required for general implementation was more training than is normally done or could be expected in most obstetric units [73]. Implementation failures can also come from inadequate resources, application to wrong populations, and failure to appreciate how cultural differences affect acceptance of new technologies.

All of the above lead to the desire/need to create “turn-key” methods. Such systems are designed to reduce or virtually eliminate the ramp-up period needed to achieve high level competence. Experience has shown sadly that without special emphasis, effort, and ongoing feedback such documented competence may not ever emerge [75, 76].

Regardless of the reasons for the failure of the above-mentioned systems to launch, attention was then turned to developing methods with automated FHR interpretation to compensate for the inability of human providers to reliably predict current fetal status so that more appropriate actions could be taken when needed. The use of AI in aiding this process must be considered and evaluated with caution [76]. “Garbage in, Garbage out” needs to be remembered [77]. While the use of AI in improving FHR-triggered alerts might make earlier intervention more likely, it would be essential to determine whether in practice the accuracy of assessing fetal condition without reference to other clinical factors would lead to better perinatal outcomes and reduce the likelihood of fetal damage.

As is the case with teratogenicity, there also needs to be some recognition of the pathophysiology of the process that makes sense. Just because “A” and “B” segregate together does not mean that one has a causal relationship with the other [78, 79]. They can be both independent results of some different process. As medicine’s use of AI increases, there could be a tendency to blithely accept the word of the “AI wizard” and possibly miss the real pathophysiology [78, 79].

#### **Automated CTG: developmental and operational issues**

**Program Development** All automated EFM programs must be able to handle online raw FHR and UC data with minimal time lag. Among the tasks required are:

- Determination of valid FHR baseline (exclusion of periodic and episodic events, management of signal loss, accounting for abrupt baseline “shifts”)
- Calculation of baseline variability (averaging R–R interval changes over predetermined time intervals)
- Identification of events (accelerations, decelerations)

- Detection of valid uterine contractions (leading edge, peak, end)
- Signal loss recognition

Key elements are to perform signal acquisition and management at the highest possible rate so that FHR pattern changes and/or significant events are recognized, and the time delay between their appearance and notation by the system is minimal. Depending on the events that are occurring, such as decelerations, there may be delays of one or more minutes following these events to ascertain that their classification is correct. Delays of recognition of significant abnormalities in FHR tracings could be disastrous both clinically and programmatically.

**Program Validation** The “gold standard” for program validation has been correlation of visual assessment by experts. This is for practical purposes generally fool-proof for baseline determination, accelerations, and some periodic events (variable, late decelerations). Such correlations are less successful, however, for variability, unless long-term variability is being considered since it is the assessment of FHR cyclic oscillation amplitude rather than “beat-to-beat.” In the latter, visual assessment is limited to crude categorization, such as “absent, minimal, or moderate.”

**Pattern Assessment** Since classification systems have been authored for FHR patterns, it would be possible to assemble the various FHR components into one or more of those that have been accepted. However, this is a complex task since such considerations as initiation, duration, and transition of a pattern’s category have degrees of arbitrariness. Decisions on these points that are incorporated into either rules or algorithms for automated EFM run the risk of arriving at inaccurate or inconsistent results. For this reason, most automated systems focus on individual features rather than overall patterns. AI could potentially be useful to combine vast amounts of data.

**Alerts** This is the most challenging aspect of automated EFM analysis. When looking at the number of FHR feature alterations that are possible (e.g., accelerations absent or present, FHR variability thresholds, decelerations classification), it is possible to create alerts that will cover many possibilities. During the course of a normal labor, most intrapartum EFM tracings will intermittently exhibit any number of these alterations. Therefore, a balance must be struck between a situation of either calling too many alerts or a situation where too few alerts are triggered. In the former case, staff will be distracted, may cease paying attention altogether, and thus run the risk of missing a critical event. In the latter case, critical events may not be highlighted on a timely basis (this concept is exactly analogous

to the discussion of the trade-off of sensitivity/specificity vs predictive values earlier in this paper) [4]. Striking this balance requires an evidence-based hierarchy of FHR features, ranging from least likely to be associated with fetal compromise to most likely to be associated with fetal compromise which has been understood for 50 years (Fig. 5). Therefore, any such alerting system must be able to demonstrate the evidence for its decision to assign specific FHR features to corresponding alert tiers.

**Decision Prompts** From their inception, basic notifications have been included in most automated systems. Examples would be signal loss (prompt: readjust equipment) or paper out (prompt: replace paper). More advanced prompts require justification, such as tachycardia (prompt: check for maternal fever, increase IV hydration), or even more advanced (recurrent late decelerations; prompt: increase IV hydration, administer maternal oxygen, discontinue oxytocin, place mother in lateral position). While advanced prompts may be linked to established protocols, no current automated EFM system typically offers them to clinicians. Rather, clinicians must assemble the EFM data and make their own management decisions in part “dumbing down” the advantages of the computerized system.

**Medicolegal Issues** A concern that must be addressed is the medicolegal status of automated EFM analysis. Any time a human function is assigned to a computerized system, it is with the assumption that there has been robust testing of this transition with comparisons of system output versus human counterparts. PeriCalm and OmniView appear to have done this, but it remains to be seen how well such systems would hold up under daily use in other obstetric units.

## Summary of Automated Systems

A number of automated systems for EFM analysis have been developed and studied for clinical implementation on labor units around the world. There are strengths and weaknesses of such systems. However, none of the systems nor the most expert of clinical interpretations have been able to identify fetal risk reliably and early enough to solve the basic problem of preventing in utero neurologic injury. Thus, we saw a clear need to rethink the problem from scratch. Our approach has been to formally include other risk factors to create a contextualized approach hoping to achieve improved statistical and clinical performance. Such initiatives have the best opportunity to enable future systems to achieve the next level of risk analysis and to offer a higher degree of protection for the fetus.

## Developing a New Approach

Even with all of the innovative approaches that have been just described, not enough has actually improved in the performance of EFM since its introduction into widespread practice in the 1970s. The most likely conclusion of that is that the inherent variance within an EFM tracing by itself does not adequately assess or predict the overall fetal condition. Analogous to outcomes from using other multimodal screening methods such as those employed for Down syndrome, we hypothesized that a contextualized approach might demonstrate better statistical performance metrics for intrapartum fetal assessment. We also recognized that by the time EFM patterns become strongly associated with intrapartum fetal asphyxia, that event has likely already occurred. The salient concept is the goal of prevention of injury by application of FHR pattern associations should enable earlier recognition of risk and enable the initiation of “intrauterine rescue” of the fetus—either by improving the intrauterine environment or by prompt delivery.

Thus, we developed a new approach for the interpretation of EFM [4, 21–30]. Our risk scoring system formally includes both known antepartum and intrapartum risk factors that can contribute to adverse neonatal neurological outcomes. We defined a new term, the “Fetal Reserve Index” (FRI). In its initial version, it is an evenly weighted calculation of 4 components of the EFM (heart rate, variability, accelerations, and deceleration) plus the presence of increased uterine activity (IUA), and various maternal, obstetrical, and fetal risk (MOFR) factors (Table 3) [4, 21–30]. 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. We have used standard ACOG definitions for everything, except that we define IUA as >4 contractions per 10 min rather than >5. We have explained these issues in detail elsewhere [3, 19–28].

The FRI was initially calculated per each 20-min segment of monitoring. With widespread computerization, it will be continuous or divided into shorter segments. In the calculation, each of 8 categories is assigned a score of “1” if the category is deemed normal and “0” if it is considered abnormal (Table 3, Fig. 7). The MOFR variables are generally 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 is calculated for the number of points divided by 8 and multiplied by 100 to give a percentage. All 8 categories

**Table 3** Fetal Reserve Index components

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%	

being normal would produce 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 is defined as ≤ 25 (corresponding to the “red zone”) (Fig. 7).

Entering the red zone occurs in about 25% of patients; it is not a call for immediate delivery. Rather, it is a cause for expeditious assessment by senior staff, who can evaluate the situation. Intrauterine resuscitation (IR) efforts should usually be the first course of action, such as stopping intravenous (IV) oxytocin infusion, repositioning the patient, increasing IV fluids, and maternal administration of oxygen

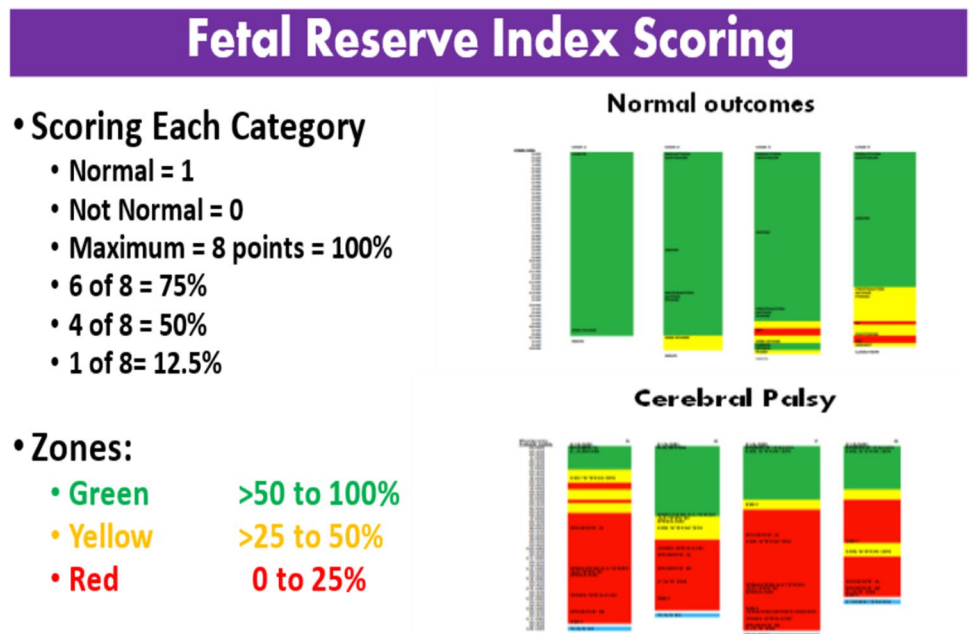
by mask. Entering the red zone also starts a “shot clock” (as is used in basketball). Our management protocol is to allow up to 40 min to get out of the red zone. Failure to do so would then start a 30-min-to-delivery protocol, as per the ACOG guidelines. Obviously, some cases could require immediate delivery (e.g., prolonged unremitting bradycardia or prolapsed cord) [4, 21–30].

Our first publication directly compared postnatal ACOG monograph criteria, CAT III criteria, and FRI in a dataset of 60 singleton term babies who entered labor with CAT I tracings and who developed CP [21]. Even in retrospect, none of these babies had any other apparent causes of their neurological compromise beyond labor issues. These infants were compared with 200 controls with normal outcomes.

Apgar scores of the CP cases were much lower. So were their cord blood pH measurements which averaged 7.03. Only 27% of the CP cases had a pH < 7.00. These results were in contradistinction to the rigidity of the ACOG monograph criteria which used a cut point of < 7.00 to consider labor-related issues to be potentially causative of CP [29]. The FRI showed significantly lower scores for CP babies than controls. Twenty-two percent of controls reached the red zone, and they were there for an average of 1 h. In contrast, CP babies “turned red” often early in labor, and they stayed in the red zone for an average of over 5 h (Fig. 6). All CP cases were “red” for at least 2 h unless a sudden “sentinel event” occurred (e.g., prolapsed cord, sudden bradycardia) in which case, the “shot clock” protocol would have ensured patients were delivered well before the 2-h 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 Monograph criteria,

**Fig. 7** Eight components of FRI. Each is weighted equally. Scores are then divided into color zones. Upper right shows 4 representative control cases. Bottom right has 4 representative cases that entered labor as Category I and went home with a baby with cerebral palsy. Each case progresses from 12 o’clock down to 6. CP cases go RED early and tend to stay there for hours



assuming we could know prenatally what could only be determined postnatally, was 28%. CAT III had a sensitivity of 45%, while the FRI had a sensitivity of 100%. While the FRI sensitivity will, of course, never stay at 100%, it was significantly better than that produced by current assessment methods [31, 32]. We now have 11 published studies with over 2000 control patients and continue to show that the FRI has far better performance metrics than the CAT system [3, 19–28]. Overall, a meta-analysis approach of our publications and combined database shows the FRI strongly outperformed CAT III (nearly 100% vs 45%) [4, 21–30].

We then expanded our focus beyond the basic outcome measures. While the prediction and prevention of fetal neurological injury are of utmost importance, the incidence of emergent operative deliveries (EODs) per se is much higher and takes its own toll on patients, families, and the entire labor and delivery staff [24]. Such interventions also have higher complication rates even when everything turns out well. Our data show the FRI can anticipate the development of need for EODs if management is not altered. Those needing EOD spent an average of 1 h 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.

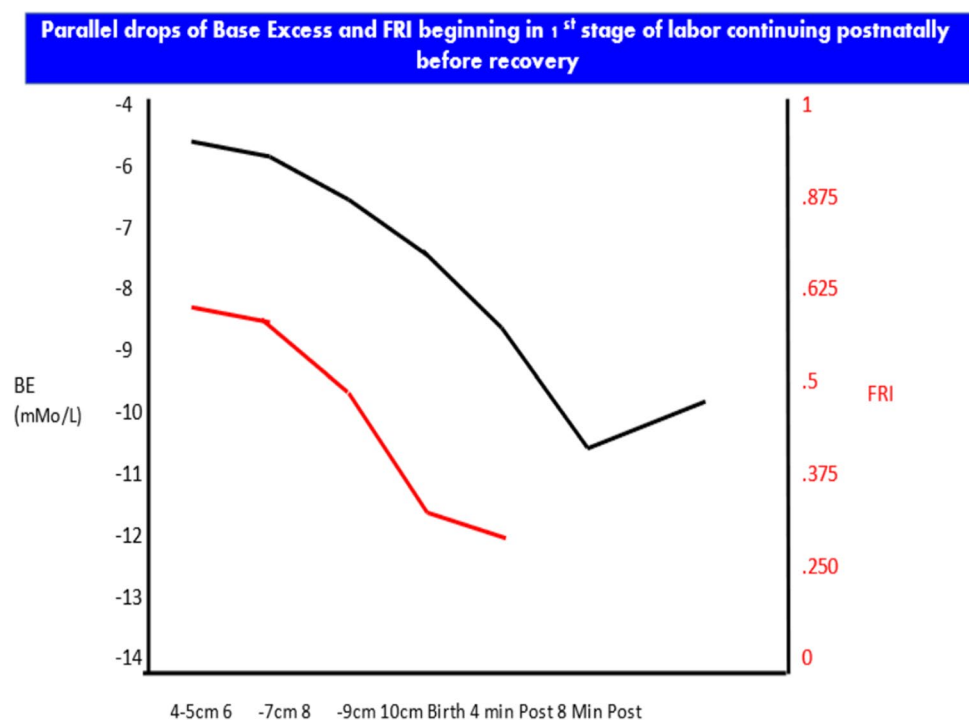
We then performed a prospective study. In 400 cases, usual clinical management was conducted. Then, we used the principles of the FRI (by expert non-computerized) management; the rate of emergency deliveries was reduced from 17 to 4% (65%), emergency cesarean deliveries decreased from 8.5 to 3.3% (62%) as the utilization of IR more than doubled (20% to 47%) [24]. These findings suggest that one

of the principal benefits of the FRI is earlier identification of problems that have a higher likelihood of being ameliorated by earlier attention.

In our next set of studies, we used a 45-year-old research database of cases managed with extensive use of FBS and continued heart rate monitoring and umbilical catheterization with arterial blood samples over the first neonatal hour [27–30]. We introduced further contextualizing the FRI score and scalp sample BE and pH by multiples of the median (MoM) for the cervical dilatation in the 1st stage of labor when the blood was obtained (same as maternal serum alpha fetoprotein by gestational age weeks). There were very different implications for same BE when viewed in the context of cervical dilatation. For example, a BE of  $-9$  mmoL/L at 9 cm is essentially at the median for that dilatation, but the same result at 4 cm would be 2 MoM and demonstrate a much higher risk profile [29] (Fig. 8). The earlier detection of deteriorating acid–base balance may permit earlier and less drastic “course correction.” Using IR in the first stage of labor dramatically reduces the chances of requiring EOD.

The second aspect of these case studies investigated the decades’ long issue concerning the use and risks of forceps deliveries. In the 1970s, Friedman et al. and others published several papers showing that the IQ of babies delivered by midforceps was lower for that of babies from spontaneous vaginal deliveries [80, 81]. In the 80 s, we challenged their conclusions [82]. After nearly 40 years later, our data showed that the pH and BE in cord blood were, in fact, lower for both low and midforceps babies. However, the FBS data showed that the differences were already there 1 h before

**Fig. 8** Deterioration of fetal acid–base balance begins early in the 1st stage of labor, continues through delivery and the first several minutes postpartum, and only after several minutes begins to recover



delivery such that, if anything, the compromise caused the forceps—not the other way around [28] (Fig. 9).

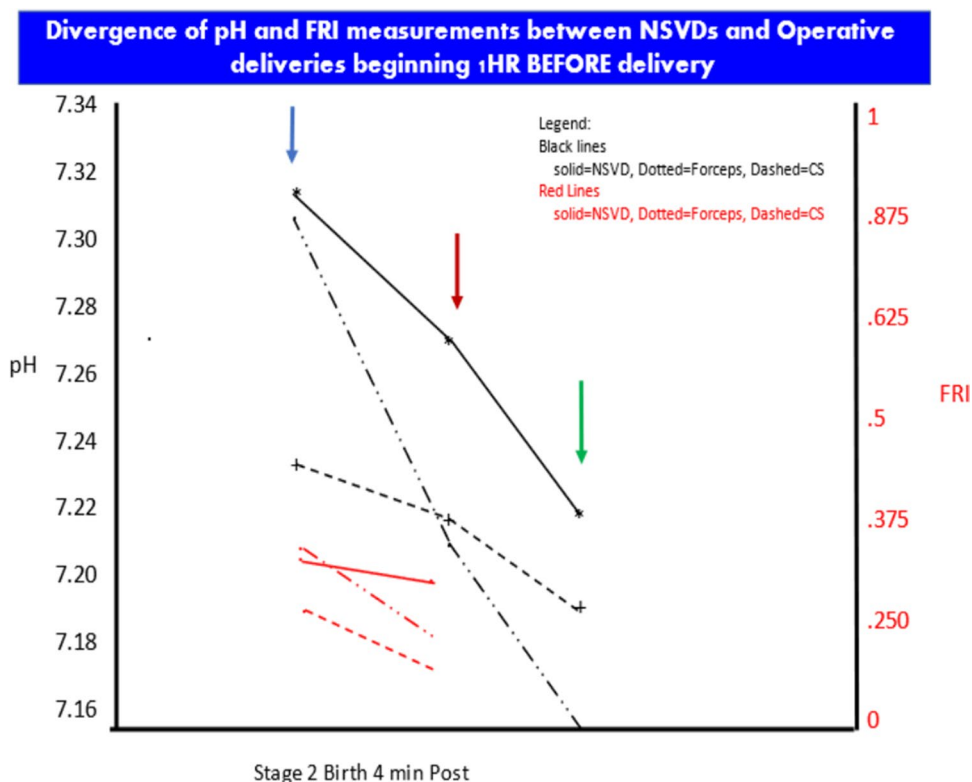
Finally, analysis of the neonatal data showed that we really need to rethink the whole process of labor, delivery, and immediate neonatal care [27]. Following delivery, about 85% of fetuses have a significant tachycardia which resolves over the first hour. Dividing patients by the last FRI score before delivery into 3 groups, the worst FRI group (about 5% of patients) had a tachycardia (average of 185 beats per minute) with loss of variability. The neonatal heart rates remain abnormal for an average of 50 min. If we assumed that the first 10 min of the neonatal tracing were the last 10 min of the fetal tracing, 25% of cases would be considered Category III (Fig. 10). At the same time, the BE routinely worsens before it gets better, and 33% of cases of dip below  $-12$  mMol/L which is considered the threshold of risk for neurological compromise (Fig. 11). Our data suggest the need to reconceptualize the process of delivery to include continued monitoring of the neonate for at least a half hour to allow earlier intervention such as brain cooling for at-risk babies [27]. Ezquerro et al. have studied the association of fetal reperfusion time and correlation to neonatal acidemia and that it works better than existing approaches. They further concluded as we have, that adding more parameters are still required for reliable interpretation of fetal well-being [83].

### Implications and Expectations

Our studies suggest the FRI provides a more reliable approach 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, particularly when CAT II tracings occur, is much too complex for front line management. There are too many variables to be informally considered, and there is no clear, straightforward method of management. Anecdotally, some experienced fetal medicine specialists have responded to the performance of the FRI by stating that they do not need such an approach because they have always factored in “other factors” in their EFM interpretations. Unfortunately, most physicians and midwives cannot reliably render such expert subjective judgements needed to overcome the limitations of the CAT system. A good analogy here is the diagnosis of myocardial infarction. For decades, the diagnosis was a construct that incorporated clinical signs and symptoms, interpretation of the ECG, and nonspecific 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 [84].

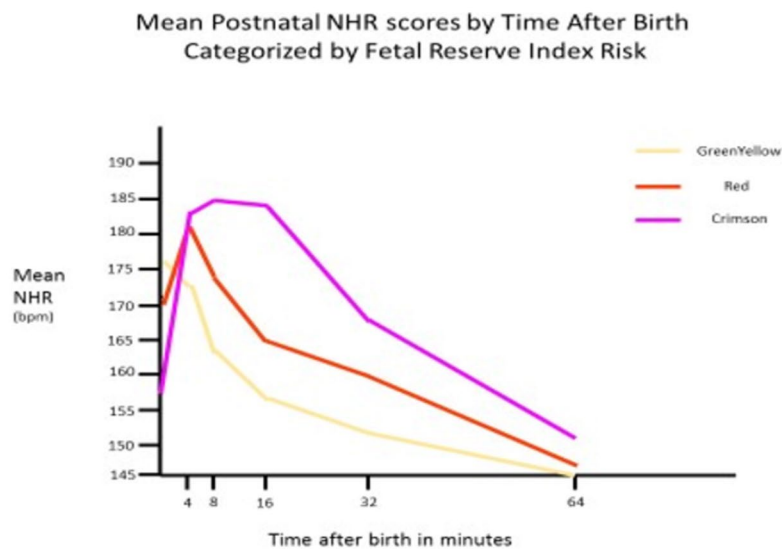
In developing the FRI, attention has been paid to each of these issues, the most important of which, we believe, is the notion of the role of EFM in avoiding fetal or maternal harm, which includes the need to avoid performing an emergency

**Fig. 9** Patients delivered by midforceps have lower pH and base excess compared with normal spontaneous vaginal deliveries. However, the differences were already there at least 1 h before delivery. The data suggest fetal compromise caused the forceps, not the other way around



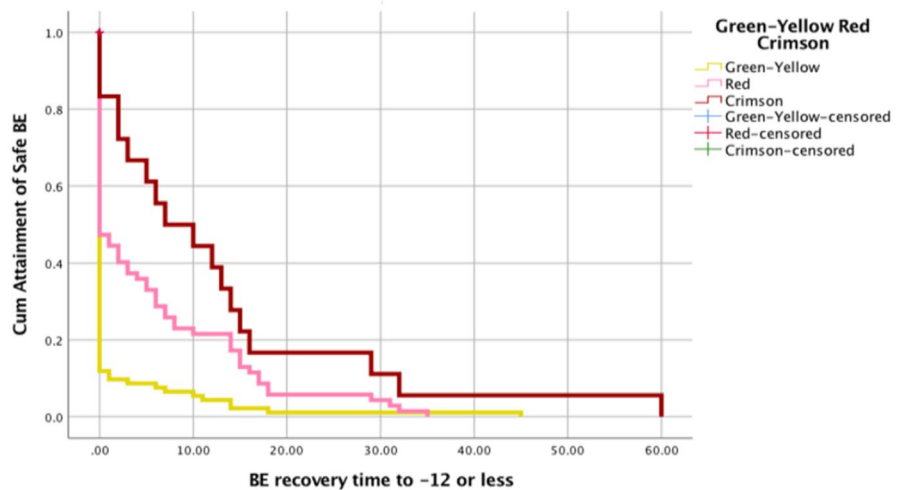
**Fig. 10** After birth, neonatal heart rate does not immediately go down. Rather, it rises for several minutes first. By dividing patients by their last FRI score before delivery, the worst group has tachycardia up to 185 bpm. If the first 10 min of the neonatal tracing were the last 10 min of the fetal tracing, 25% of cases would be considered Category III

**85 % OF NEONATES HAVE TACHYCARDIA: Extent correlates with FRI score**



**Fig. 11** Base excess recovery time over 1st hour postpartum. Again, dividing by last FRI score, the worst group has over 10 min of acidemia in the “at risk range” of greater than or equal to  $-12$  mMol/L for neurologic impairment

**BE RECOVERY TIME CORRELATES WITH FRI SCORE**



delivery during labor [27]. We have attempted to change the objective of surveillance from trying to decide the severity of asphyxia and need to “rescue” the fetus to “keeping the fetus out of harm’s way in the first place.” We do this by switching our focus on diagnosis of the severity of acidemia to recognizing the in utero status in which the fetus cannot be guaranteed to be normal but is not yet likely damaged. As such, contextualized 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 of the fetus, the red zone is often 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. The former generally allows more time than the latter for IR to attempt to halt the progression of such deterioration.

We treat the EFM tracing as a language, albeit still an imperfect one to be sure. We use this “language” to query the fetus, not asking, “What is your pH? but, rather asking “How did you tolerate 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 on admission from those deemed to be abnormal on admission. Our approaches in our published studies have focused on how behaviorally normal neurologically intact fetuses respond to the stresses and events of labor. If on admission to the labor unit, the fetus is determined to be already compromised, then a different set of management approaches apply (not discussed here).

Our studies to date have progressed considerably through the phase of development towards establishing proof of principle. Automation is underway to make our approach practical for frontline use and will be followed by large-scale studies using data in an electronic medical records format. The final stage of development will require live implementation.

We consider EFM as only a screening test, not a diagnostic test. 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 that has only been demonstrated for the STAN system [73, 74].

Generalizable concepts from our approach suggest the need to see the “big picture” first and then cone down to specific circumstances. Nearly two decades ago, we suggested that the protocols of tertiary referrals have been generally backwards (i.e., a less-trained provider is deciding if a patient should be triaged upwards) [85, 86]. 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 Nicolaidis et al. later suggested for prenatal care [87]. 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.

## Moving Forward

In many fields, AI is held out as the savior just around the corner [88]. It is still very early to assess the role of AI for the prediction of intermediate and end point outcomes in Obstetrics. For the FRI, our “back of the napkin” version already appears to offer better prediction of intrapartum fetal status than does the Category system. There certainly are opportunities to improve several aspects of FRI even further. It is parallel to the often-used distinction in the modeling literature between measurement models (how you measure things) and substantive models (how you explain things) [88] and to the idea of three kinds of continual re-specification: what things are (and are not), how they are related (and not related), and what should be included in a model (or left out). It is clear that other things beyond standard EFM interpretation are needed if EFM is to be improved. Thinking outside the EFM box implies expanding the models currently used to include both antenatal and intrapartum information regarding levels of risk that might have contextual implications for the analysis of EFM patterns. AI can contribute both in terms of improving our understanding of how to measure both EFM patterns and other pertinent variables. Ultimately, AI might improve the capacity to alert physicians to problematic situations that are being commonly missed and that can have potentially deleterious consequences.

## Contextualized Systems

There seem to be at least four dimensions that are important for understanding the nature of the directions in which automated systems are changing and what will be needed for them to become more widely adopted [87]. Examples of this process include (1) the level of sophistication of AI when incorporated in the system designs, (2) the degree of contextualization of clinical factors that may influence pregnancy outcomes, and (3) the ability to predict the impact of these factors early enough to potentially alter the course of labor management; and finally, (4) the clinical usability and interpretability of the information generated by these systems.

All of the systems that have been evaluated to date and have varying deficiencies. Oxford Sonicaid is not intended as an intrapartum system. OB TraceVue is well established for issuing alerts, but it does not provide specific guidance for interventions. STAN has been in use for two decades as a decision support system, using an intervention matrix based on FHR patterns and ECG changes, that has also been extensively studied. While it has shown some benefits

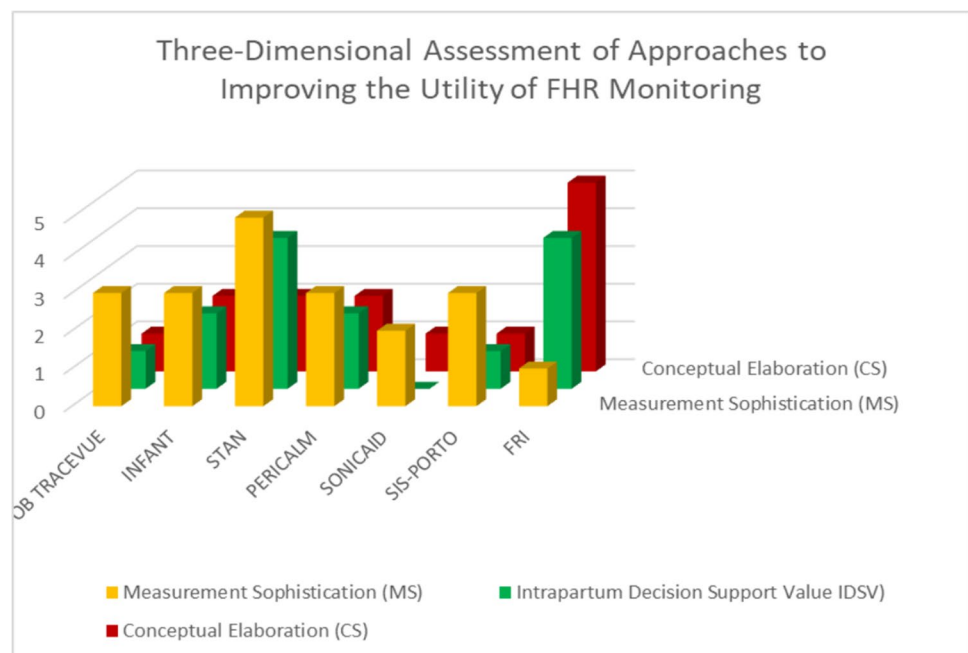
in reducing metabolic acidosis, there is a long learning curve, and it is difficult for some users to adopt into their clinical practices. SIS-PORTO is a proprietary FHR interpretative system that has recently been used in conjunction with STAN, but it has shown only questionable benefits to date. INFANT is a hybrid system for FHR analysis and risk assessment with AI programming, but it has not been shown to be beneficial. It also has a lengthy learning curve. The FRI has documented the utility of contextualizing FHR information, and through a number of small studies has been shown to improve screening metrics for a variety of outcomes and offers the prospect of becoming alerted to possible interventions earlier in labor. Its use of AI is still in the early training stage. Finally, PeriCALM Plus with iCue™ is a hybrid system that employs AI for FHR and uterine activity analysis. It is capable of displaying trends in FHR tracings and labor progress, but no large trials have been conducted to show benefit.

To facilitate discussion of three of the dimensions of automated FHR analytic systems, so as to represent these scores in three-dimensional space (Fig. 12), leaving aside the issue of interpretability and the conflated aspects of the length of the learning curve. These scores are an approximate first stab, but they allow us to see rather clearly the fundamental problem with most approaches: the clinically used systems have focused almost exclusively just on FHR factors. There has only been token, if any, consideration for the *context* within which the extent of abnormality and trends in abnormality of such characteristics over time take place. As such, they all cluster in the same part of a three-dimensional space.

In demonstrating this clustering, crude though it might be, we can suggest what needs to happen if work in this area is to move forward. The most fundamental problem seems to be the lack of consideration of additional and readily available clinical factors beyond those of the various FHR features and patterns. It seems apparent that the significance of worsening clusters of FHR features might vary according to the prior risk established for the mother and fetus. So, one dimension of change in the application of EFM needs to be the systematic, theory-driven incorporation of more information in these systems, with AI potentially playing a major role. Secondly, over time, with the incorporation of AI, more refined measures of FHR characteristics, and their implications in different risk contexts, will be identified, possibly increasing the ability to make more timely clinical interventions. Thirdly, much more attention needs to be given to the usefulness of alerts, including how early they are given, how salient they are, and how clear are the options for intervention. Lastly, beyond the equipment, attention needs to be given to the development of viable and useful clinical management systems.

Going forward, it is clear that there must be additional information incorporated into any automated system that uses EFM interpretation as a part of overall intrapartum management. Establishing risk of fetal compromise at the beginning of labor will require input of preexisting maternal and fetal factors, examination of findings such as cervical examination and fetal station, and medications currently being administered (e.g., anesthetic agents, oxytocin). The degree of detail must be malleable so that there should be little or no controversy about what is being recorded. Such steps as how often risk

**Fig. 12** Our modeling of 3 dimensions of multiple systems employed for interpretation of EFM. Each of the 7 systems we reviewed was categorized for (a) Measurement sophistication, (b) Conceptual elaboration, and (c) Intrapartum support value. We scored each as being low level (0) up to (5) as the most



status is reassessed will depend on the starting points and what changes subsequently appear. A decision tree, if that is the form of support to be taken, will need to be evidence-based. To achieve this end and since we want to query the system before adverse outcomes happen, AI will likely be invoked since what is really being asked for is a likelihood of fetal compromise rather than an absolute diagnosis of fetal compromise. In fact, established AI systems like those used to predict weather trends or storm behavior generate probabilities rather than absolute outcomes [89, 90].

## Conclusions

EFM has evolved over the past 50 years, but it has not produced the massive reduction in neurologic injury and CP that had been anticipated at its inception. There are multiple reasons for EFM's failures as we have detailed here. Perhaps the most fundamental ones are a lack of clarity of what EFM could do, imprecision in its measurements coupled with considerable variability in its interpretation, and widely varying applications to intrapartum management. We have attempted to put these issues into perspective and to show how a different approach might improve intrapartum fetal assessment even while using the same basic measurements that are already being collected. The application of contextualized management approaches to EFM has the potential to improve intrapartum care both for infants and mothers. Only then could we realize the dream of better outcomes that was put forward at the inception of this technology.

**Author Contribution** Drs. Evans, Britt, Devoe, and Ms. Evans all participated in the evaluation of data, preparation of the manuscript, and its editing. The authors read and approved the final manuscript.

**Funding** None.

**Data Availability** All previously published.

**Code Availability** N/A.

## Declarations

**Ethics Approval** N/A

**Consent to Participate** N/A

**Consent for Publication** N/A

**Conflicts of Interest** Dr. Evans has patents on the Fetal Reserve Index.

## References

- <https://medium.com/swlh/everything-that-can-be-invented-has-been-invented-49c4376f548b#:~:text=In%201889%2C%20Charles%20H.,be%20invented%20has%20been%20invented.%E2%80%9D>
- Evans MI. Overcoming militant mediocrity. *Am J Obstet Gynecol.* 2008;198:656–61.
- Riopelle, K, Gluesing, J, Alcordo, T, Baba, M, Britt, D, McKether, W., Monplaisir, L, Ratner, H, and Wagner, KH “Context, task and the evolution of technology use in global virtual teams.” 2003 Chapter 11 in CB Gibson and SG Cohen (Eds) *Virtual Teams that Work: Creating Conditions for Virtual Team Effectiveness.* New York: Jossey-Bass.
- Evans MI, Eden RD, Britt DW, Evans SM, Schiffrin BS. Reconceptualizing fetal monitoring. *Eur J Gynecol Obstet.* 2019;1:10–7.
- 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(15):878–81.
- Bloom SL, Spong CY, Varner MW, Rouse DJ, Weininger S, et al. Fetal pulse oximetry and cesarean delivery. *N Engl J Med.* 2006;355:2195–202.
- Low LK, Chuey M, Abfel Nabi S, De Vries R. Explaining the (over)use of electronic fetal monitoring from a study of work flow on the unit. *Obstet Gynecol.* 2020;135:1605.
- Spector-Bagdady K, De Vries R, Hope-Harris L, Kane Low L. Stemming the standard of care SPRWL: clinician self-interest and the case of electronic fetal monitoring. *Hastings Center Reports.* 2017;47:16–24.
- Cohen A, Hanft R. Policy directions for effective evaluation and management. In: Hanft ABCaRS, editor. *Technology in American Health Care.* Ann Arbor, MI: University of Michigan Press; 2004. p. 480.
- Evans M, Hanft R. The introduction of new technologies. *ACOG Clinical Seminars.* Washington: ACOG; 1997. p. 3.
- Good Reads. <https://www.goodreads.com/quotes/4079-a-new-scientific-truth-does-not-triumph-by-convincing-its> (Accessed 0502–2021)
- Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva: World Health Organization; 1968.
- Galen R, Gambino R. Beyond normality: Wiley biomedical publication; 1975. p237
- Iilria Emin E, Emin E, Papalois A, Willmott F, Clarke S, Sideris M. Artificial intelligence in obstetrics and gynaecology. *In Vivo.* 2019;33:1547–51.
- Wikipedia. Precision and Recall. ([https://en.Wikipedia.org/wiki/Precision\\_and\\_Recall](https://en.Wikipedia.org/wiki/Precision_and_Recall)) Accessed 05–02–2021
- Evans MI, Evans SM, Bennett TA, Wapner RJ. The price of abandoning diagnostic testing for cell-free fetal DNA screening. *Prenat Diagn.* 2018;38(4):243–5.
- 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(10):730–4.
- 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(7):734–42.
- 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(13):1172–7.
- Grimes DA, Peipert JF. Electronic fetal monitoring as a public health screening program: the arithmetic of failure. *Obstet Gynecol.* 2010;116(6):1397–400.
- Eden RD, Evans MI, Evans SM, Schiffrin BS. The “fetal reserve index”: re-engineering the interpretation and responses to fetal heart rate patterns. *Fetal Diagn Ther.* 2018;43:90–104.
- Eden RD, Evans MI, Evans SM, Schiffrin 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.
23. 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;21:1–7.
  24. 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” using the fetal reduction index. *J Matern Fet Neonat Med.* 2019;32:2561–9.
  25. Evans MI, Britt DW, Eden RD, Gallagher P, Evans SM, 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;26:858–63.
  26. 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 Neonat Med.* 2018:1–9.
  27. Eden RD, Evans MI, Britt DW, Evans SM, Gallagher P, Schifrin BS. Combined prenatal and postnatal prediction of early neonatal compromise. *J Matern Fet Neo Med.* 2021. <https://doi.org/10.1080/14767058.2019.1676714>.
  28. Evans MI, Britt DW, Evans SM: Midforceps didn't cause “compromised babies” – “compromise” caused forceps: an approach towards safely lowering the cesarean delivery rate. *J Matern Fet Neo Med (in press)* <https://doi.org/10.1080/14767058.2021.1876657>
  29. Evans MI, Britt DW, Eden RD, Evans SM, Schifrin BS: Earlier and improved screening for impending fetal compromise. *J Matern Fet Neo Med (in press)* <https://doi.org/10.1080/14767058.2020.1811670>
  30. Evans MI, Britt DW, Worth J, Mussalli G, Evans SM, Devoe LD. Uterine contraction frequency in the last hour of labor: how many contractions are too many? *J Matern Fetal Neonat Med (in press)*
  31. ACOG. Neonatal encephalopathy and cerebral palsy: defining the pathogenesis and pathophysiology. Washington, DC: ACOG; 2003 January, 2003. 94 p.
  32. ACOG. Neonatal encephalopathy and neurologic outcome. D'Alton M, editor: ACOG; 2014.
  33. Jorgensen JS, Weber T. Fetal scalp blood sampling – a review. *ACTA Obstet Gynecol.* 2014;93:548–55.
  34. Clark SL, Nageotte MP, Garite TJ, Freeman RK, Miller DA, Simpson KR, et al. Intrapartum management of category II fetal heart rate tracings: towards standardization of care. *Am J Obstet Gyn.* 2013;209:89–97.
  35. 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.* 2016.
  36. Intrapartum fetal heart rate monitoring: nomenclature, interpretation, and general management principles ACOG Practice Bulletin No. 70; December 2005. (superseded by Practice Bulletin No. 106, July 2009).
  37. Sartwelle TP, Johnston JC. Continuous electronic fetal monitoring during labor: a critique and a reply to contemporary proponents. *Surg J.* 2018;4:e23–8.
  38. Evans MI, Britt DW: Categorization of cerebral palsy cases. *Am J Obstet Gynecol* 2021;225:210-211
  39. Nelson KB, Ellenberg JH. Antecedents of cerebral palsy. I. Univariate analysis of risks. *Am J Dis Child.* 1985;60:1031–8.
  40. Nelson KB, Ellenberg JH. Antecedents of cerebral palsy. *NEJM.* 1986;315:81–6.
  41. Ellenberg JH, Nelson KB. The association of cerebral palsy with birth asphyxia: a definitional quagmire. *Dev Med Child Neurol.* 2013;55:210–6.
  42. Wapner RJ, Martin CL, Levy B, et al. Chromosomal microarray versus karyotype for prenatal diagnosis. *N Engl J Med.* 2012;367:976–85.
  43. Jelin AC, Vora N. Whole exome sequencing: applications in prenatal genetics. *Obstet Gynecol Clin North Am.* 2018;45:69–81.
  44. Nakao M, Okumara A, Hasegawa J, Toyokawa S, Ichizuka K, Kanayama N, et al. Fetal heart rate pattern in term or near term cerebral palsy: a nationwide cohort study. *Am J Obstet Gynecol.* 2020;223(907):e1-13.
  45. Moreno-DeLuca A, Millan F, Pesacreta DR, Elloumi HZ, Oetjens MT, Teigen C, et al. Molecular diagnostic yield of exome sequencing in patients with cerebral palsy. *JAMA.* 2021;325:467–75.
  46. Kurten C, Knippel A, Verde P, Kozlowski P. A Bayesian risk analysis for trisomy 21 in isolated choroid plexus cyst: combining a prenatal database with meta-analysis. *J Matern Fetal Neonat Med.* 2021;34:889–97.
  47. Rossi AC, Prefumo F. Antepartum and intrapartum risk factors for neonatal hypoxic -ischemic encephalopathy: a systematic review with meta-analysis. *Curr Opin Obstet Gynecol.* 2019;31:410–7.
  48. Graham EM, Ruis KA, Hartman AL, et al. A systematic review of the role of intrapartum hypoxia-ischemia in the causation of neonatal encephalopathy. *Am J Obstet Gynecol.* 2008;199:587–95.
  49. Evans K, Rigby AS, Hamilton P, et al. The relationship between neonatal encephalopathy and cerebral palsy: a cohort study. *J Obstet Gynaecol.* 2001;21:114–20.
  50. Jonsson M, Agren J, Norden-Lindeberg S, et al. Neonatal encephalopathy and the association to asphyxia in labor. *Am J Obstet Gynecol.* 2014;211(667):e1-667/e8.
  51. Rei M, Ayres-de-Campos D, Bernardes J. Neurological damage arising from intrapartum hypoxia/acidosis. *Best Pract Res Clin Obstet Gynaecol.* 2016;30:79–86.
  52. 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. *Obstet Gynecol.* 2008;112:661–5.
  53. Ayres-de-Camos D, Spong CY, Chandraran E. FIGO consensus guidelines on intrapartum fetal monitoring cardiotocography. *Int J Gynaecol Obstet.* 2015;131:13–24.
  54. Bulletin 2016; pp 3–5
  55. Cahill AG. Umbilical artery pH and base deficit in obstetrics. *Am J Obstet Gynecol.* 2015;213:257–8.
  56. Hamilton EF, Warick PA. New perspectives in electronic fetal surveillance. *J Perinat Med.* 2013;41:83–90.
  57. Elliot C, Warick PA, Graham E, Hamilton EF. Graded classification of fetal heart rate tracings: association with neonatal metabolic acidosis and neurologic morbidity. *Am J Obstet Gynecol.* 2010;202(258):e1-8.
  58. Nelson KB, Sartwelle TP, Rouse DJ. Electronic fetal monitoring and caesarean section: assumptions versus evidence. *BMJ.* 2016;355:i6405.
  59. Alfirevic Z, Gyte, GML, Cuthbert A, Devane D. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. *Cochrane Database Syst Rev.* 2017; (2):CD006066. <https://doi.org/10.1002/14651858.CD006006.pub3>.
  60. Alonso-Betanzos A, Moret-Bonillo V, Devoe L, Searle J, Banias B, Ramos E. Computerized antenatal assessment: the NST-Expert Project. *Automedica.* 1992;14:3–22.
  61. Devoe L, Golde S, Kilman Y, Morton D, Shea K, Waller J. A comparison of visual analyses of intrapartum EFM according to the then new National Institute of Child Health and Human Development guidelines for computer analyses by an automated FHR monitoring system. *Am J Obstet Gynecol.* 2000;183(2):361–6.

62. The Infant Collaborative Group. Computerized interpretation of fetal heart rate during labor (INFANT): a randomized controlled trial. *Lancet*. 2017;389:1719–29.
63. Ayres-de-Campos D, Bernardes J, Garrido A, Marques-de-Sá J, Pereira-Leite L. SisPorto 2.0: a program for automated analysis of cardiotocograms. *J Matern Fetal Med*. 2000;9(5):311–8.
64. FM-ALERT: a randomised clinical trial of intrapartum fetal monitoring with computer analysis and alerts versus previously available monitoring. Inês Nunes, Diogo Ayres-de-Campos, Austin Ugwumadu, Pina Amin, Philip Banfield, Antony Nicoll, Simon Cunningham, Paulo Sousa, Cristina Costa-Santos, João Bernardes, for the FM-ALERT study group. [www.ecic2015.org](http://www.ecic2015.org)
65. Elliott C, Warrick PA, Graham E, Hamilton EF. Graded classification of fetal heart rate tracings: association with neonatal metabolic acidosis and neurologic morbidity. *Am J Obstet Gynecol*. 2010;202(3):258.e1–8.
66. Bloom SL, Spong CY, Thom E, Varner MW, Rouse DJ, Weininger S, et al. Fetal pulse oximetry and cesarean delivery. *N Engl J Med*. 2006;355(21):2195–202.
67. Amer-Wählin I, Bordahl P, Eikeland, et al. ST analysis of the fetal electrocardiogram during labor: Nordic observational multicenter study. *J Mat Fet Neonat Med*. 2001;12:260–6.
68. Westgate J, Harris M, Cumow JSH, Greene KR. Plymouth randomized trial of cardiotocogram only vs ST waveform plus cardiotocogram for intrapartum monitoring in 2400 cases. *Am J Obstet Gynecol*. 1993;169:1151–60.
69. Amer-Wählin I, Hellsten C, Norén H, Hagberg H, Herbst A, Kjellmer I, Lilja H, Lindoff C, Månsson M, Mårtensson L, Olofsson P, Sundström AK, Maršál K. Cardiotocography only versus cardiotocography plus ST analysis of fetal electrocardiogram for intrapartum fetal monitoring: a Swedish randomised controlled trial. *Lancet*. 2001;358:534–8.
70. Per Olofsson, Diogo Ayres-de-Campos, Jörg Kessler, Britta Tendal, Branka M Yli, Lawrence Devoe. A critical appraisal of the evidence for using cardiotocography plus ECG ST interval analysis for fetal surveillance in labor. Part I: the randomized controlled trials. *Acta Obstet Gynecol Scand*. 2014;93(6):556–68.
71. Per Olofsson, Diogo Ayres-de-Campos, Jörg Kessler, Britta Tendal, Branka M Yli, Lawrence Devoe. A critical appraisal of the evidence for using cardiotocography plus ECG ST interval analysis for fetal surveillance in labor. Part II: the meta-analyses. *Acta Obstet Gynecol Scand*. 2014;93(6):571–86.
72. Belfort MA, Saade GR, Thom E, Blackwell SC, Reddy UM, et al. A randomized trial of intrapartum fetal ECG ST-segment analysis. *N Engl J Med*. 2015;373(7):632–41.
73. Landman A, Immink-Duijker ST, Mulder EJH, Koster MPH, Xodo S, Visser GHA, Groenendall F, Kwee A. Significant reduction in umbilical artery metabolic acidosis after implementation of intrapartum STS waveform analysis of the fetal electrocardiogram. *Am J Obstet Gynecol*. 2019;221(1):63.e1–63.e13.
74. Devoe LD, Ross M, Wilde C, Beal M, Lysikewicz A, Maier J, Vines V, Amer-Wählin I, Lilja H, Noren H, Maulik D. United States multicenter clinical usage of the STAN 21 electronic fetal monitoring system. *Am J Obstet Gynecol*. 2006;195(3):729–34.
75. Schultes MT, Ajjaz M, Klug J, Fixsen DL. Competencies for implementation science: what trainees need to learn and where to learn it. *Adv Health Sci Ed*. 2021;26:19–35.
76. Peterson HB, Haidar J, Fixsen D, Ramaswamy R, Weiner BJ, Leatherman S. Implementing innovations in global women's, children's and adolescent's health: realizing the potential for implementation science. *Obstet Gynecol*. 2018;131:423–420.
77. Evans MI, Krantz D, Hallahan T, Sherwin J. Impact of NT credentialing by FMF, NTQR, or both upon screening distribution and performance. *Ultrasound Obstet Gynecol*. 2012;39:181–4.
78. Gao S, He L, Chen Y, Li D, Lai K. Public perception of artificial intelligence in medical care: content analysis of social medial. *J Med Internet Res*. 2020;22:e16649.
79. Calado AM, Dos Anjos PM. An overview of teratogenicity. *Methods Mol Biol*. 2018;1797:3–32.
80. Friedman EA, Sachtleben MR. High risk labor. *J Reprod Med*. 1971;7:52–7.
81. Friedman EA, Sachtleben MR, Bresky PA. Dysfunctional labor XII. Long term effects on infant. *Am J Obstet Gynecol*. 1977;127:779–83.
82. Richardson DA, Evans MI, Cibils LA. Midforceps: a critical review. *Am J Obstet Gynecol*. 1983;145:621–32.
83. Ezquerro MC, Cornudella RS, Esteban LM, Zamora del Poso, C, Espiau Romera A, Castan Larraz, B Castan Mateo S: Total intrapartum fetal reperfusion time (fetal resilience) and neonatal acidemia. *J Matern Fetal Neonat Med* (in press). 2021. <https://doi.org/10.1080/147670581915977>.
84. Galen RS, Reiffel JA, Gambino R. Diagnosis of acute myocardial infarction. Relative efficiency of serum enzyme and isoenzyme measurements. *JAMA*. 1975;232(2):145–7.
85. 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(4):253–8.
86. Britt DW, Eden RD, Evans MI. Matching risk and resources in high-risk pregnancies. *J Matern Fetal Neonatal Med*. 2006;19(10):645–50.
87. Nicolaides KH. Turning the pyramid of prenatal care. *Fetal Diagn Ther*. 2011;29(3):183–96.
88. SAS. [https://www.sas.com/id\\_id/insights/articles/analytics/machine-learning-and-artificial-intelligence-in-a-brave-new-world.html?utm\\_source=google&utm\\_medium=cpc&utm\\_campaign=ai-ml-us&utm\\_content=GMS101515&gclid=Cj0KCQjwI6KFBhCgARIsALB0g8KoRtRMeHfqaYa5ug3AUTiV16ZzuaMTd7sDQTNmjB-42\\_ZWUmwwcVMaAhHiEALw\\_wcB](https://www.sas.com/id_id/insights/articles/analytics/machine-learning-and-artificial-intelligence-in-a-brave-new-world.html?utm_source=google&utm_medium=cpc&utm_campaign=ai-ml-us&utm_content=GMS101515&gclid=Cj0KCQjwI6KFBhCgARIsALB0g8KoRtRMeHfqaYa5ug3AUTiV16ZzuaMTd7sDQTNmjB-42_ZWUmwwcVMaAhHiEALw_wcB) (accessed 5 22 21)
89. Britt DW. A conceptual introduction to modeling: qualitative and quantitative perspectives. Mahwah, NJ: Lawrence Erlbaum Associates; 1997.
90. Alcorn T. How AI can make weather forecasting less cloudy. *Wall Street Journal* April 4, 2021. <https://www.wsj.com/articles/how-ai-can-make-weather-forecasting-less-cloudy-11617566400>