Bayes Theorem and Protopathic Bias: Methodological Concerns When Addressing the Impact of Fetal Heart Rate Patterns on the Cesarean Section Rate

Jacques Balayla, MD, MPH, CIP, FRCSC[1]   Ariane Lasry, BSc[1]   Yaron Gil, MD[1]
Alexander Volodarsky-Perel, MD[1]

[1] Department of Obstetrics and Gynecology, McGill University, Montreal, Quebec, Canada

Am J Perinatol Rep 2020;10:e342–e345.

Address for correspondence Jacques Balayla, MD, MPH, CIP, FRCSC, Department of Obstetrics and Gynecology, Jewish General Hospital, McGill University, 3755 Chemin de la Côte-Sainte-Catherine, Montréal, QC, Canada H3T 1E2 (e-mail: Jacques.balayla@mail.mcgill.ca).

Abstract

Over the last 30 years, the caesarean section rate has reached global epidemic proportions. This trend is driven by multiple factors, an important one of which is the use and inconsistent interpretation of the electronic fetal monitoring (EFM) system. Despite its introduction in the 1960s, the EFM has not definitively improved neonatal outcomes, yet it has since significantly contributed to a seven-fold increase in the caesarean section rate. As we attempt to reduce the caesarean rates in the developed world, we should consider focusing on areas that have garnered little attention in the literature, such as physician sensitization to the poor predictive power of the EFM and the research method biases that are involved in studying the abnormal heart rate patterns—umbilical cord pH relationship. Herein, we apply Bayes theorem to different clinical scenarios to illustrate the poor predictive power of the EFM, as well as shed light on the principle of protopathic bias, which affects the classification of research outcomes among studies addressing the effects of the EFM on caesarean rates. We propose and discuss potential solutions to the aforementioned considerations, which include the re-examination of guidelines with which we interpret fetal heart rate patterns and the development of noninvasive technologies that evaluate fetal pH in real time.

Keywords
► fetal pH
► caesarean section
► Bayes theorem
► protopathic bias

Cesarean sections are the most commonly performed surgical procedure worldwide, with an average one-third of the babies in the developed world being born via cesarean delivery.[1] While cesarean delivery can be lifesaving, its widespread use over the past 30 years has been associated with significant maternal and neonatal morbidity, without substantial obstetrical or neonatal benefits beyond a rate of 10 to 20%.[2,3] As such, international efforts have aimed to control cesarean rates through clinical, administrative, educational, and advocacy means.[4] Likewise, international organizations like FIGO (International Federation of Gynecology and Obstetrics) and the WHO (World Health Organization) have deployed taskforces to address this problem.[1,5]

The primary causes for the increase in cesarean delivery rates observed over the last three decades are multifactorial but can be generally categorized into financial/administrative (higher income for providers, predictable scheduling, and personnel costs for elective cesareans), legal (perceived decreased risk of litigation), and technical reasons (increase in use of fetal monitoring, increased use of assisted reproductive technologies and multiple gestations, increased prevalence of maternal comorbidities, among others).[6] Likewise, impediments to the reduction of the cesarean section rate are also numerous and complex, and revolve around the aforementioned reasons.

When these causes are studied in epidemiological reports, several methodological concerns, which are seldom

addressed and accounted for, often limit the research efforts. In the case of the electronic fetal monitoring (EFM) in particular, these have to do with its intrinsic ability to diagnose fetal distress and how we classify and interpret outcomes when the fetal monitor is used. It is reasonable to hypothesize that raising awareness about these may bring about important changes that may impact not only the way we conduct and interpret research reports, but in the way we care for patients and frame our clinical judgment as well.

The Electronic Fetal Heart Rate Monitor

Perhaps no other clinical factor has had a greater disproportionate influence into the cesarean section rate than the introduction of the electronic fetal heart rate (FHR) monitor (EFM) in the 1960s.[7] These authors believe therein lies ample potential for intervention to reduce this rising trend.

Adequate fetal and neonatal development depends upon the presence of a normal acid–base environment during pregnancy and the smooth transition from intrauterine to extrauterine life. Current methods to assess fetal pH and acid–base status are invasive and carry significant maternal and fetal risks. Given these limitations, obstetrical care providers developed the EFM system, a noninvasive tool which evaluates baseline variability in FHR patterns to evaluate fetal wellbeing/distress in real-time.[7] The concept of fetal distress is a comprehensive, umbrella term. Though several risk factors have been described to characterize fetal distress (meconium staining, maternal fever, abnormal fetal tracings, among others), we refer herein to cases of hypoxia/acidosis leading to a low fetal pH at birth.

Though there is ample physiological evidence that FHR patterns are inextricably linked to fetal acid–base status, the use of EFM has not been shown to reliably predict neonatal pH, nor has it reduced the incidence of adverse perinatal outcomes as initially marketed, including long-term neurological morbidity and cerebral palsy. On the other hand, the introduction of the EFM has contributed to an increase in the cesarean section rate six to sevenfold without any significant benefit.[8] Despite the physiological basis for the EFM patterns, the poor specificity associated with the current interpretation of the EFM therefore leads to a paradox that has been coined as the "Obstetrical Paradox."[7] To better and more accurately estimate the predictive ability of the EFM, we need to consider Bayes' theorem in this clinical context.

Bayes' Theorem

Bayes' Theorem describes the probability of an event occurring based on prior knowledge of conditions that might be related to that specific event. Mathematically speaking, the equation translates to the conditional probability of an event A given the presence of an event or state B. Simply stated, the equation is written as follows:

$$P(A|B) = P(B|A) \times P(A)/P(B)$$

where,

A = event A

B = event B

$P(A|B)$ = probability of A given B is true.

$P(B|A)$ = probability of B given A is true.

$P(A)$ and $P(B)$ are the independent probabilities of A and B.

In the context of FHR tracings and abnormal fetal pH, we define event A as the abnormal fetal pH and event B as the abnormal tracing. Therefore, the equation asks: given an abnormal tracing, what is the probability of having an abnormal fetal pH?

P (abnormal pH | abnormal tracing) =

P (abnormal pH | abnormal tracing) × P (abnormal pH)/ P (abnormal tracing)

We could expand Bayes theorem using proper epidemiological terminology. The probability of testing positive when you have the disease is the probability of a true positive test divided by the probability of all (true and false) positive tests. Therefore, the probability of having an abnormal pH (A) in setting of an abnormal tracing (B) is:

$$P(A|B) = P(B|A) \times P(A)/P(B|A) \times P(A) + P(B|{\sim}A) \times P({\sim}A)$$

$P(B|A)$ = true positive rate or sensitivity.

$P(A)$ = prevalence of abnormal pH in the study population.

$P(B|{\sim}A)$ = false positive rate or 1-specificity.

$P({\sim}A) = 1 - P(A)$

Using the latter definition, we can better estimate the risk of abnormal cord acid–base status in the presence of abnormal tracings. Cahill et al[9] addressed the EFM–abnormal pH relationship by studying its sensitivity and specificity in different clinical scenarios classified either by formal categories or specific tracing characteristics as defined by the NICHD (National Institute of Child Health and Human Development). Taking the values from Cahill et al,[9] and the prevalence of acidemia in the population of 1.7%, we can use Bayes theorem to determine the diagnostic accuracy of the EFM (►**Table 1**).

In essence, what the Bayes theorem equation yields in this scenario is the positive predictive value (PPV) of the EFM, which is dependent on the population prevalence of the condition in question—in this case $P(A)$, an abnormal fetal pH.[10] We observe that for an average prevalence of fetal acidemia between 1.5 and 2% in the population, the predictive ability of a fetal tracing is lower than 5% in several commonly encountered clinical scenarios.

Despite its widespread clinical use, the poor predictive value of the EFM is well-documented.[11] Indeed, the major value of the EFM lies primarily in its negative predictive value —where a normal tracing is predictive of a normal acid–base status in the newborn in over 95% of the cases.[12]

Having said that, Bayes' theorem provides the clinician with a way to think critically in terms of conditional probabilities—what is colloquially known as "Bayesian thinking." Simply put, Bayesian thinking stipulates that the risk of an event taking place is not only defined by the specific clinical context of that particular case, but also, by how common that event occurs in the general population. Even in the presence of what is suspected to be overwhelming evidence of fetal compromise in the fetal monitor, the very low overall prevalence of fetal hypoxia influences the predictive accuracy of the monitor interpretation.

Table 1 Sensitivities and specificities of EFM in different scenarios[9]

EFM features discriminative of acidemia (umbilical cord pH)	Applying Bayes theorem
P(A) = prevalence of acidemia = 1.7%	

• Tracing: always category II
  ○ Sensitivity 67.8%
  ○ Specificity 50.2%
	• 2.30% chance of an abnormal pH in setting of an always category II tracing.

• Tracing: Ever category III
  ○ Sensitivity: 69.1%
  ○ Specificity: 50.0%
	• 2.33% chance of an abnormal pH in setting of an ever category III tracing.

• Tracing: Ever moderate variability
  ○ Sensitivity 73.8%
  ○ Specificity 44.3%
	• 2.24% chance of an abnormal pH in setting of ever moderate variability.

• Total number of decelerations
  ○ Sensitivity 68.5%
  ○ Specificity 58.8%
	• 2.61% chance of an abnormal pH in setting of total decelerations.

• Composite total deceleration area, ever tachycardia and mostly moderate variability
  ○ Sensitivity 64.6%
  ○ Specificity 78.5%
	• 4.94% chance of an abnormal pH in setting of composite total deceleration area, ever tachycardia + mostly moderate variability.

Abbreviation: EFM, electronic fetal monitoring.

Like Ashby and Smith, we argue that the natural framework for evidence-based medicine should be a Bayesian approach to decision-making that incorporates an integrated summary of the available evidence, reliability of assessment tools, and clinical data.[13] Unlike a *Bayesian* view, which assigns a probability to a hypothesis, a *frequentist* view tests a hypothesis without assigning a probability; in the frequentist inference, conclusions are drawn from sample data by emphasizing the frequency or proportion of the data—a type of statistical inference.[14] In our scenario, a frequentist approach thus translates to the direct performance of a caesarean to test the hypothesis about the validity of the abnormal tracing as it pertains to an abnormal pH, without considering crucial conditions related to the event in question, notably, the prevalence of an abnormal cord pH in the population. It is our contention that a more consistent application of a Bayesian analytic approach to the abnormal heart rate patterns—umbilical cord pH relationship would reduce the EFM contribution to the caesarean rate as it would shed light on its very poor predictive ability.

Protopathic Bias

Consider a fetus in labor that displays an abnormal tracing. The baby is urgently delivered, usually after intrauterine resuscitation is performed in an attempt to improve fetal

oxygenation (e.g., stopping oxytocics, changing maternal positions, giving maternal fluids, stimulating the fetal scalp, among others). If the arterial cord pH and base excess were normal at birth, we could draw two equally valid conclusions. We could consider the interpretation of the tracing as a "false positive": the tracing was suspicious for fetal compromise, but the oxygenation status of the newborn was actually normal and unaffected by the delivery method. However, we could also perform an emergency cesarean and conclude that the normal acid–base status observed in the newborn was conferred by the emergent cesarean, thus associating the intervention to the prevention of injury and to a normal pH. The latter is a far more concerning scenario, because it reinforces the false belief that cesarean sections are saving all newborn lives—prompting undoubtedly, should we adopt this opinion, an increase in the cesarean rate for abnormal tracings. This is an example of protopathic bias—a type of misclassification bias.

Protopathic bias arises when the initiation of a drug, treatment, or intervention occurs in response to an early manifestation/symptom/sign of a disease under study that is not yet diagnosed. Acting on that early manifestation may lead to inappropriate conclusions about the causal relationship between the intervention and outcome, should that outcome eventually occur. For example, the use of analgesics in response to pain caused by an undiagnosed tumor might lead to the erroneous conclusion that the analgesic caused the tumor. Likewise, protopathic bias can lend credence to the opposite effect—namely, that the intervention led to an improvement in outcomes when no such effect took place. The latter may take place in a considerable proportion of cases where a caesarean section, the intervention, is undertaken for an abnormal tracing in a fetus whose acid–base status is actually normal. The erroneous attribution of benefit to the intervention can furthermore lead to a differential misclassification of outcomes in research studies where the efficacy of the intervention is praised. In the case of the EFM–pH relationship, a differential misclassification bias therefore occurs when the normal pH is erroneously attributed to the caesarean delivery, which may in fact not have played any role in more than nine out of ten cases, since we have shown the PPV of the EFM to be <5% in several common clinical scenarios. Protopathic bias may overemphasize the risk in the exposed group with abnormal tracing, and consequently overestimate the benefit of emergent caesarean delivery for the prevention of abnormal cord pH.

The Solutions

Given the aforementioned concerns and the fact that EFM interpretation remains one of the most likely venues to tackle the cesarean epidemic, what solutions are necessary to address the limitations of the EFM and its contribution to this issue? First, sensitization of obstetrical providers to Bayesian analysis and the poor PPV of FHR patterns is key and may impact clinical decision-making in the acute setting. However, in our view, tackling the "obstetrical paradox"—the inherent limitation of the EFM as defined

in this manuscript—is a more fundamental and critical task. The latter will undoubtedly require a re-examination of the guidelines we use to interpret the EFM to improve its sensitivity, specificity, and predictive values. Though attempts to measure fetal oxygenation directly using pulse oximetry and ST segment analysis have been shown to be of little value in monitoring the fetus during labor, perhaps no step will address their limitation more effectively than the development of noninvasive technologies that can evaluate fetal pH—not merely oxygen saturation—in real time, thus minimizing any uncertainty about the ability of the EFM to predict true fetal distress.[15] Realistically, so long as we do not have a noninvasive diagnostic test for fetal pH, the above considerations may not significantly change the clinical management of patients with category II or III tracings. After all, though the risk of abnormal pH in different clinical scenarios (PPV) is very small, there is no feasible way to ascertain fetal pH in labor until after delivery. To this end, we are currently developing the FETAL technique (Fourier evaluation of tracings and acidosis in labor)—a tool that would evaluate fetal acid–base status noninvasively and in real-time during labor.[15]

Conclusion

As increasing cesarean deliveries continue to be a major source of maternal and neonatal morbidity, analyzing the individual root causes may allow for discovery of strategies not yet employed in reducing the cesarean rate. In particular, a critical review of the EFM—a major contributor to the cesarean section epidemic—through the lens of Bayesian analysis and protopathic bias, as well as the development of technologies to assess fetal cord pH in real time is a necessary step toward significantly improving outcomes for mothers and babies alike.

Conflict of Interest
The authors have no conflicts of interest to report.

References

1 Betran AP, Torloni MR, Zhang JJ, Gülmezoglu AM; WHO Working Group on Caesarean Section. WHO statement on caesarean section rates. BJOG 2016;123(05):667–670

2 Burrows LJ, Meyn LA, Weber AM. Maternal morbidity associated with vaginal versus cesarean delivery. Obstet Gynecol 2004;103(5 Pt 1):907–912

3 De Luca R, Boulvain M, Irion O, Berner M, Pfister RE. Incidence of early neonatal mortality and morbidity after late-preterm and term cesarean delivery. Pediatrics 2009;123(06):e1064–e1071

4 Sakala C, Mayberry LJ. Vaginal or cesarean birth? Application of an advocacy organization-driven research translation model. Nurs Res 2006;55(Suppl 2):S68–S74

5 Visser GHA, Ayres-de-Campos D, Barnea ER, et al. FIGO position paper: how to stop the caesarean section epidemic. Lancet 2018;392(10155):1286–1287

6 Porreco RP, Thorp JA. The cesarean birth epidemic: trends, causes, and solutions. Am J Obstet Gynecol 1996;175(02):369–374

7 Balayla J, Shrem G. Use of artificial intelligence (AI) in the interpretation of intrapartum fetal heart rate (FHR) tracings: a systematic review and meta-analysis. Arch Gynecol Obstet 2019;300(01):7–14

8 Alfirevic Z, Devane D, Gyte GM, Cuthbert A. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. Cochrane Database Syst Rev 2017;2:CD006066

9 Cahill AG, Tuuli MG, Stout MJ, López JD, Macones GA. A prospective cohort study of fetal heart rate monitoring: deceleration area is predictive of fetal acidemia. Am J Obstet Gynecol 2018;218(05):523.e1–523.e12

10 Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Stat Med 1997;16(09):981–991

11 Grimes DA, Peipert JF. Electronic fetal monitoring as a public health screening program: the arithmetic of failure. Obstet Gynecol 2010;116(06):1397–1400

12 Vintzileos AM, Nochimson DJ, Antsaklis A, Varvarigos I, Guzman ER, Knuppel RA. Comparison of intrapartum electronic fetal heart rate monitoring versus intermittent auscultation in detecting fetal acidemia at birth. Am J Obstet Gynecol 1995;173(04):1021–1024

13 Ashby D, Smith AF. Evidence-based medicine as Bayesian decision-making. Stat Med 2000;19(23):3291–3305

14 Coory MD, Wills RA, Barnett AG. Bayesian versus frequentist statistical inference for investigating a one-off cancer cluster reported to a health department. BMC Med Res Methodol 2009;9:30

15 Balayla J, Shrem G. Solving the obstetrical paradox: the FETAL technique—a step toward noninvasive evaluation of fetal pH. J Pregnancy 2020;2020:7801039