Prediction of preterm birth in symptomatic women using decision tree modeling for biomarkers

Jacquelyn L. Hill, MSc; M. Karen Campbell, PhD; Guang Yong Zou, PhD; John R. G. Challis, PhD; Gregor Reid, PhD; Hiroshi Chisaka, MD; Alan D. Bocking, MD

OBJECTIVE: The objective of the study was to use recursive partitioning (RP) to identify gestational age–specific and threshold values for infectious and endocrine biomarkers of imminent delivery.

STUDY DESIGN: RP was developed using a previously collected data set and then applied to a prospectively collected cohort of women in threatened preterm labor. Predictors of preterm birth were considered, including white blood cell count (WBC), corticotrophin-releasing hormone (CRH), cortisol, and maternal age.

RESULTS: At 22–27 weeks’ gestation, WBC of greater than 12,000/mL was the most accurate predictor of delivery within 48 hours; at 28–31 weeks’ gestation, CRH of greater than 684 pg/mL was the most accurate predictor; and at 32–26 weeks’ gestation, CRH and maternal age were the most important variables.

CONCLUSIONS: These results indicate that maternal WBC greater than 12,000/mL prior to 28 weeks’ gestation and CRH beyond 28 weeks are the most accurate biomarkers in predicting preterm birth within 48 hours. RP assists in establishing clinically relevant and gestational age–specific threshold levels for these variables.

Key words: corticotrophin-releasing hormone, infection, preterm labor, recursive partitioning


P reterm birth continues to be a major challenge in obstetrics, and there are few methods available to reliably predict true preterm labor in women who present with symptoms of labor. Currently transcervical ultrasound measurements and/or cervicovaginal fetal fibronectin levels are the most commonly used diagnostic tools. These techniques both have high negative predictive values but relatively low positive predictive values. Consequently, many women and their fetuses are exposed unnecessarily to tocolytic drugs and corticosteroids and are admitted to hospital. These admissions and treatments create emotional and financial stress on the mother and family as well as significant financial costs to society. We and others have previously shown that maternal corticotropin-releasing hormone (CRH) levels are higher in symptomatic women who give birth within 48 hours, compared with those who do not.

Our previous prospective study involving 218 symptomatic women showed that, at different gestational ages, different factors were associated with preterm birth within 48 hours. White blood cell count (WBC), a marker of subclinical infection, was the only variable shown to be predictive at 22–27 weeks, supporting previously published work suggestive of the gestational age dependence (particularly less than 32 weeks) of the relationship between markers of infection and preterm birth. Beyond 28 weeks’ gestation, CRH and adrenocorticotropic hormone were predictors of birth within 48 hours, indicating an early activation of the fetal hypothalamic-pituitary-adrenal axis. Additionally, at 32–36 weeks’ gestation, demographic and lifestyle variables were significantly associated with preterm birth.

For the purposes of clinical decision making, it is useful to identify threshold levels of biomarkers, above which preterm birth is probable. Recursive partitioning (RP) to produce decision trees has been increasingly utilized to identify accurate biomarker cutoffs for obstetrical conditions including preterm birth. Prediction rules derived from decision tree analysis and based on critical thresholds of variables are simpler and more clinically applicable than the results of more conventional analyses. For example, logistic regressions (LR) can treat predictive variables as continuous or categorical and can estimate effects with statistical accuracy, but the interpretation of parameters are more abstract and do not as readily provide decision-making thresholds. However, RP analysis has limitations, including the potential for spurious findings because the findings are highly depen-
dent on the data set used in the analysis. We know of no previous reports describing the use of classification trees to predict preterm birth among symptomatic women with the biomarkers of interest in this study.

The objectives of this study were: (1) to explore, by RP, the relationship of gestational age–specific endocrine, infection, and other variables with delivery within 48 hours using our previously published data, (2) to compare results derived by LR and RP; and (3) in a new prospective cohort of symptomatic pregnant women, to apply the threshold values for infectious and endocrine biomarkers developed in the previous cohort using RP. We were also interested in the potential utility of assessment of cervicovaginal fetal fibronectin measurements, in these women because at the time of this study, there was minimal information regarding its predictive value for preterm birth in symptomatic women.

Materials and Methods

This study was approved by the University of Western Ontario Review Board for Research Involving Human Subjects. Briefly, RP was developed using a previously collected data set and then applied to a prospectively collected cohort termed validation study.

Derivation data and analysis

Data for 201 of the 218 women in our earlier study, excluding 17 for missing data, were partitioned with respect to preterm birth within 48 hours of admission or not, by the algorithm of Atkinson and Therneau. implemented in S Plus 6.1 (Seattle, WA). The resulting classification tree graphically summarizes the predictors, in order of their discriminating power relative to each other. The method also selects the optimum threshold level for each predictor. The highest-risk subgroups are easily identified in the tree by the combinations of predictors. Analysis assigned equal misclassification costs for false positives as for false negatives and used the Gini index for splitting nodes.

For all 3 gestational age categories, the following predictor variables were considered: WBC, cortisol, cervical dilatation 2-4 cm, CRH, multiparous, smoker, maternal age, contractions at admission, and cervical effacement of at least 50%. Maternal age, endocrine variables, and WBC were continuous scale variables, whereas cervical dilatation, parity, smoking, contractions, and cervical effacement were binary variables.

In addition to developing the RP algorithm, for 32–36 weeks’ gestation only, sample size permitted the further evaluation of the use of these cutpoints in a LR analysis.

Validation data

Two hundred seven women with singleton pregnancies, who presented with threatened preterm labor to St Joseph’s Health Care London, a tertiary perinatal center in London, Canada, from April 1999 to July 2002, were recruited to the study after obtaining informed consent. Sample size was based on our previous study with an anticipated birth rate within 48 hours of presentation of 30%. Exclusions were multifetal gestations, fetal anomalies, maternal diabetes mellitus, abruptio placenta, preeclampsia, intrauterine growth restriction, cervical dilatation greater than 4 cm, ruptured membranes, and clinical signs of infection (WBC greater than 18,000/mL).

A sterile speculum examination was performed in all women prior to digital examination and cervicovaginal swabs were taken to determine the presence of bacterial vaginosis (BV) by Nugent scoring. In women who were eligible, cervicovaginal swabs were taken for determination of fetal fibronectin levels using a Dacron swab. Swabs were immediately placed in buffer provided by the manufacturer (Adzea Biomedical Corp, Sunnyvale, CA). Measurements were made using a commercially available enzyme-linked immunosorbent assay (Adzea Biomedical Corp). Results were considered positive at 50 ng/mL or greater.

Maternal blood samples were obtained at the time of recruitment, and WBC was determined from whole blood before centrifugation to collect plasma. Aliquots of plasma were immediately frozen and stored at -80°C for later determination of CRH and cortisol by radioimmunoassay. Intra- and interassay coefficients of variation for CRH previously measured were 12% and 6.5%, respectively. Interassay coefficients of variation for cortisol measurements were approximately 11%. CRH and cortisol values were normalized using gestational age appropriate norms from an earlier study of subjects who gave birth at term using the Standardization Procedure of SAS (version 8.02, Cary, NC). Gestational age was determined by a combination of last menstrual period and early ultrasound examination.

Results

Derivation study

Using the derivation data at 22–27 weeks’ gestational age, RP analysis indicated that WBC was the most accurate predictor of preterm delivery within 48 hours (Figure 1). A threshold of the 48th percentile of these data (12,000/mL) was important and, for those with WBC 12,000/mL or greater, further differentiation was provided by the next most accurate predictor, cortisol 55th percentile or greater (202 ng/mL). Fourteen women at this gestational age exceeded both thresholds and, of these, 10 (71%) delivered within 48 hours. In contrast, none of the 21 women with WBC less than 12,000/mL delivered within 48 hours. Competitive variables for the first predictive split in the tree, in lieu of WBC, were cortisol and cervical dilatation. At higher levels of WBC, competitive variables for the second split, in lieu of cortisol, were maternal age and WBC (at a different cutpoint).

For women of 28–31 weeks’ gestational age, CRH was the most accurate predictor of birth within 48 hours. Of 8 women exceeding the CRH threshold of the 83rd percentile (684 pg/mL), 5 (63%) gave birth within 48 hours as compared with 4 of the 43 (9%) of women with CRH less than 684 pg/mL (Figure 2). Competitive variables, in lieu of CRH, were WBC, cortisol, cervical effacement, and maternal age.

For gestational ages 32–36 weeks (Figure 3), CRH was the most predictive variable with a first split at a threshold
value of the 61st percentile for this gestational age category (1368 pg/mL). For those with CRH less than 1368 pg/mL, maternal age (30 years or older) was the next splitting variable. For those with CRH 1368 pg/mL or greater, cortisol 18th percentile or greater (152 ng/mL) was the next splitting variable. As indicated in Figure 3, women with CRH levels of 1368 pg/mL or greater and cortisol levels of greater than 152 ng/mL had an 83% chance of delivering within 48 hours. Women with CRH levels of less than 1368 pg/mL but maternal age 30 years or older had a 75% chance of delivering within 48 hours. The low-risk groups were younger women with low CRH or women of any age for whom CRH, but not cortisol, was elevated.

Table 1 presents the results of LR for the derivation data, within the 32-36 weeks’ gestational age group, incorporating all 3 variables identified by recursive partitioning algorithm of Atkinson and Therneau (RPART). The independent effects of CRH of 1368 pg/mL or greater, cortisol of 152 ng/mL or greater, and maternal age of 30 years or older were to increase the odds of birth within 48 hours by factors of 5.1, 3.7, and 4.7, respectively.

Validation study
Of the 207 symptomatic women recruited to the validation study, 101 (49%) were admitted to the hospital with a diagnosis of threatened preterm labor. Forty-six (22%) received a course of glucocorticoids and 13 (6.3%) received tocolytics. Forty-eight of these women (23%) gave birth less than 37 weeks, and of those, 17 (8.2%) gave birth within 48 hours. One hundred fifty women were previously pregnant, and of these, 46 (30.7%) had a previous preterm birth. Demographic and clinical characteristics of the women enrolled are provided in Table 2. Nugent scores were available in 185 women and the prevalence of BV was 37.3%. There was no difference in Nugent score between women who gave birth within 48 hours, compared with those who did not (Table 3).

For all gestational age groups in the validation study, there is high sensitivity, good specificity, but low positive predictive values because of the low prevalence of the outcome in the data set (Table 4). At 22-27 weeks’ gestation, all of the subjects delivering within 48 hours had peripheral WBC counts of 12,000/mL or more, resulting in a highly sensitive test. None of the subjects with WBC counts less than this threshold delivered within 48 hours, resulting in very good negative predictive value.

Cortisol was not included in determining test accuracy in this gestational age group because a large proportion of records were missing cortisol data. For 28-31 weeks’ gestation, all of the women from the test set who delivered within 48 hours had CRH values above the threshold, indicating a highly sensitive test. The majority of women who did not deliver within 48 hours (85%) had CRH levels below the threshold of 684 pg/mL, indicating a highly specific predictor of preterm birth in symptomatic women. Because only 20% of women with CRH values above 684 pg/mL delivered within 48 hours, CRH had poor positive predictive value for this sample of women but had excellent negative predictive value.

Shaded boxes are terminal nodes in which the majority of subjects delivered within 48 hours. Nonshaded boxes are terminal nodes in which the majority of subjects delivered more than 48 hours after presentation.

The prediction algorithm for the 32-36 weeks’ gestational age group was applied using CRH as the primary variable with and without maternal age as a secondary variable. The sensitivity was high (80%), and when using CRH alone as a predictive variable, the specificity was high (84%). Again, the low prevalence of the outcome is responsible for the low positive predictive value.

Fetal fibronectin levels were obtained in only 104 of the women enrolled in the validation study because the remainder of women were ineligible for this test because of a recent pelvic examination or vaginal bleeding. Fetal fibronectin was positive in 8 women and negative in 96. The positive and negative predictive values for delivery within 48 hours based on a positive fetal fibronectin level were 50% and 98%, respectively.

**Comment**

These results are consistent with our previous findings that maternal CRH levels are predictive of preterm birth within 48 hours of presentation in symptomatic women at 32 weeks’ gestation and beyond. This study restricted the sample to women with intact membranes and therefore excludes a major confounder in our previous studies (ruptured membranes). We used RP for the first time to select cutpoints of biological markers to predict preterm delivery in symptomatic women.

At 22-27 weeks’ gestation, maternal blood WBC count of 12,000/mL or more, a marker of subclinical infection or inflammation, was the most effective indicator of preterm birth. Because subjects with very high WBC were already excluded from this study, this represents

---

**TABLE 1**

<table>
<thead>
<tr>
<th>Dichotomized variable</th>
<th>Variable fit</th>
<th>Odds ratio (90% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRH</td>
<td>$P = .001$</td>
<td>5.1 (2.3, 11.3)</td>
</tr>
<tr>
<td>Cortisol</td>
<td>$P = .036$</td>
<td>3.7 (1.3, 10.9)</td>
</tr>
<tr>
<td>Maternal age</td>
<td>$P = .002$</td>
<td>4.7 (2.1, 10.9)</td>
</tr>
</tbody>
</table>

Area under the curve of receiver-operating curve using the test data $= 0.77$.


**TABLE 2**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married/common law</td>
<td>184 (88.9)</td>
</tr>
<tr>
<td>Completed high school</td>
<td>170 (82.0)</td>
</tr>
<tr>
<td>Completed college/university</td>
<td>128 (62.0)</td>
</tr>
<tr>
<td>Admitted to hospital</td>
<td>101 (49.0)</td>
</tr>
<tr>
<td>Primiparous</td>
<td>57 (27.5)</td>
</tr>
<tr>
<td>Smoker</td>
<td>55 (26.8)</td>
</tr>
<tr>
<td>Glucocorticoid administration</td>
<td>46 (22.0)</td>
</tr>
<tr>
<td>Caesarian delivery</td>
<td>37 (18.0)</td>
</tr>
<tr>
<td>Treatment with tocolytics</td>
<td>13 (6.3)</td>
</tr>
</tbody>
</table>

Of the 150 multiparous women, 46 (30.7%) had a history of preterm birth.

a change within the normal range. Evidence shows that births of very young gestational age are associated with infection, both clinically and subclinically.6,21

Another study compared amniotic fluid WBC to maternal blood WBC with respect to outcome of histological chorioamnionitis in laboring women with intact membranes and thus chose a higher cutpoint (14,700/mL) to distinguish women with the outcome, based on a compromise between sensitivity and specificity.22 WBC values at this higher level or higher resulted in a significantly shorter amniocentesis to delivery interval than those women with WBC less than 14,700/mL. There is considerable evidence that there is a generalized activation of the immune system during labor consistent with these findings.

Our findings not only support previous parametric multivariable analysis but also add to this by identifying a threshold WBC value of 12000/mL, above which the probability of preterm delivery is high in this sample. This is of relevance to clinical decision making. The high negative predictive value of WBC when less than 12000/mL is reflective of the overall low prevalence of preterm birth in the sample. WBC less than 12000/mL indicates a low probability of delivering preterm and a low risk of intrauterine or maternal subclinical infection.

At 28-31 weeks’ gestation, the RP algorithm selected CRH of greater than 684 pg/mL as the most accurate predictor of birth within 48 hours. At this gestational age, a negative test result also has importance for clinical decision making. Testing of the CRH threshold in the validation set showed high diagnostic accuracy and negative predictive value among this sample of women. There were however, a high number of false positives attributable to the low prevalence of the outcome of interest. The CRH threshold of greater than 684 pg/mL identified by RP may therefore be useful as a diagnostic tool only for deciding which women should not receive tocolytic and glucocorticoid therapy at these gestational ages.

A positive test value may not be as informative. The presence of subclinical infection is also an important factor for preterm birth at 28-31 weeks’ gestational age because WBC was a competitor for the first predictive split of the RP tree, but it is not of incremental predictive value once a prediction has been generated using CRH.

At 32-36 weeks’ gestation, WBC as a marker of infection is not predictive of preterm birth, a finding consistent with our previous multivariable analyses.5 The most predictive variable was CRH with a threshold of 1368 pg/mL or greater, indicating a 70% risk of birth within 48 hours. Analysis of the derivation data indicated that prediction

---

**TABLE 3**

<table>
<thead>
<tr>
<th>BV (Nugent score)</th>
<th>All subjects (n = 207)</th>
<th>Delivery within 48 h (n = 17)</th>
<th>Delivery greater than 48 h (n = 190)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (0-3)</td>
<td>55 (29.7%)</td>
<td>4 (33.3%)</td>
<td>51 (29.5%)</td>
</tr>
<tr>
<td>Intermediate (4-6)</td>
<td>61 (33.0%)</td>
<td>3 (25.0%)</td>
<td>58 (33.5%)</td>
</tr>
<tr>
<td>BV (7-10)</td>
<td>69 (37.3%)</td>
<td>5 (41.7%)</td>
<td>64 (37.0%)</td>
</tr>
<tr>
<td>Missing</td>
<td>22</td>
<td>5</td>
<td>17</td>
</tr>
</tbody>
</table>


---

**TABLE 4**

<table>
<thead>
<tr>
<th>Gestational age at presentation</th>
<th>Delivery within 48 h (%)</th>
<th>Accuracy and predictive values of cutpoints applied to validation data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Derivation data</td>
<td>Validation data</td>
</tr>
<tr>
<td>22-27 ks³</td>
<td>12 (28.0)</td>
<td>3 (6.7)</td>
</tr>
<tr>
<td>28-31 weeks</td>
<td>9 (18.0)</td>
<td>2 (3.6)</td>
</tr>
<tr>
<td>32-36 wks⁴</td>
<td>53 (50.0)</td>
<td>12 (11.3)</td>
</tr>
<tr>
<td>32-36 wks⁵</td>
<td>80</td>
<td>84</td>
</tr>
<tr>
<td>Total with outcome</td>
<td>74 (37.0)</td>
<td>17 (8.2)</td>
</tr>
<tr>
<td>Total in set</td>
<td>201</td>
<td>207</td>
</tr>
</tbody>
</table>

Number of women from derivation data who presented with threatened preterm labor were 43, 51, and 107 at gestational ages of 22-27, 28-31, and 32-36 weeks, respectively. For the validation data, there were 45, 56, and 106 women, respectively.

PPV, positive predictive value; NPV, negative predictive value.

³ The RP algorithm was applied using only WBC as predictive variable.

⁴ The RP algorithm was applied using CRH and age as predictive variables.

⁵ The RP algorithm was applied using only CRH as predictive variables.

within this group can be further refined with 2 sets of rules. If CRH is higher than 1368 pg/mL, we can predict an 83% chance of imminent preterm birth if cortisol is above 152 ng/mL. Otherwise, the probability of preterm birth is low (13%), even though CRH was elevated. That is, elevated CRH plus elevated cortisol predict a high risk. Conversely, even when CRH is low, we cannot rule out imminent preterm birth except in women who are younger than 30 years old.

Level 2 of this decision rule was not borne out by the validation data in which CRH by itself was highly predictive, but adding a second level of decision, on the basis of maternal age, to those with lower cortisol levels, actually decreased the specificity of the test. The value of refining the prediction using cortisol could not be assessed in the validation set because of missing data. Consistently, in both the derivation and test data, CRH was highly predictive and, if borne out in future studies, provides a straightforward decision rule that would be easy to implement.

The RP algorithm in this study was derived using previously published data arising from a cohort of women, which included some with ruptured membranes. In the derivation data set, 79 of 218 women had proceeded to preterm birth within 48 hours of presentation, but of those 79 women, 57 had ruptured membranes. The RP algorithm was tested in the validation data arising from a sample that excluded women with ruptured membranes. A limitation in the validation study was that only a small percentage of the sample within 48 hours of presentation. The a priori sample size estimate had been based on the assumption that 30% of the sample would progress into full labor, but the prevalence of true preterm labor was low in this sample, which included only women with intact membranes.

Future studies in women with intact membranes will require larger sample sizes to allow for this low prevalence of true preterm labor. Additionally, because RP methods are very data driven, they can produce spurious findings. Thus, caution is advised in applying this algorithm prior to further replication and testing in other populations. Another study limitation was standardizing CRH values (for percentile purposes) within the entire range of gestational ages in each gestational age group. This limitation would apply most to the group at 32-36 weeks’ gestation because normally there is a large difference in CRH levels from 32 to 36 weeks’ gestational age.

Future studies should standardize CRH values to more precise gestational ages, thus an additional reason for requiring larger sample sizes. As a strength of the study, these results support the important role of the maternal immune system in preterm labor, consistent with our previous studies of WBC mRNA microarray in women with threatened preterm labor.23

Cervicovaginal fetal fibronectin measurements were possible in only 47% of women in this study, consistent with the fact that this was carried out in a tertiary perinatal center at which women often had a prior digital pelvic examination. This is a recognized limitation of fetal fibronectin determinations, although recent studies would support its use in regional programs in which swabs are taken prior to pelvic examination and analysis performed only if required. Although currently transvaginal ultrasound assessments of cervical length are widely used to predict preterm birth in symptomatic women, at the time of this study, this was not readily available. It would be important to perform further studies in symptomatic women combining fetal fibronectin measurements, cervical length assessments, and biochemical measurements as described in this study. In addition, there is recent interest in the application of proteomic and transcriptomic methodologies to develop new diagnostic tests in women symptomatic for preterm labor.

In this cohort of predominantly middle socioeconomic class, Caucasian women, we found that the prevalence of BV as assessed by Nugent score was 37.3%. Although this prevalence was higher than many previous studies, it was not predictive of birth within 48 hours.

This analysis has added to our previous LR analysis in a similar cohort of symptomatic women by showing hierarchies of variables predictive of preterm birth and the respective cutoff points of these markers most useful for accurate prediction of preterm delivery. In terms of a choice between the use of RP or LR, each provides a different component of analysis. The RP gives a clinically useful prediction algorithm, whereas the LR analysis provides statistically rigorous identification and estimation of the effects of predictive variables on the outcome of interest. In our data, the RP analysis has supported results of LR analysis, showing that the studied biomarkers were more predictive than clinical or demographic markers.

Most potential biomarkers and other measures of preterm birth investigated from women with threatened preterm labor are characteristically similar, with respect to diagnostic performance and accuracy, in that negative predictive values are superior to positive predictive values and the tests are usually more specific than sensitive. These results show gestational age-specific predictors and respective cutoffs of preterm delivery in 2 populations of women with threatened preterm labor, thus compartmentalizing some of the factors associated with preterm birth. By using a combination of biomarkers predictive of preterm birth to diagnose imminent preterm delivery, diagnostic accuracy, and predictive values should increase. Discovery of other predictors with high positive predictive value in combination with currently investigated biomarkers of preterm birth should also contribute to improved methods of therapy for symptomatic women.

ACKNOWLEDGMENTS
We thank the nurses and physicians in the Family Birthing Centre at St Joseph’s Health Care London for their assistance and support for this study. In addition, the contributions of Lorna Froste, Tammy Thompson, Larry Stitt, and Carole Watson to the recruitment of subjects and data analysis are gratefully acknowledged.
REFERENCES


DISCUSSION

George A. Macones, MD. I would like to thank the society for asking me to serve as a discussant for this paper, and I would like to congratulate Dr Bocking on his interesting paper and excellent presentation.

Preterm birth remains the foremost issue in perinatal medicine today. Preterm birth, defined as delivery at less than 37 completed weeks, is common, with an annual incidence of 9-15% in the US population. At high-risk centers, such as ours in St Louis, the incidence is even higher, approaching 25%. Preterm birth is important clinically, not just because of its high incidence but also because of its strong association with neonatal morbidity and mortality. In addition, preterm births place enormous stress on families and on our health care system. For these reasons, many researchers are working to understand mechanisms of preterm birth and testing interventions to reduce its occurrence.

Preterm births are generally categorized as spontaneous or indicated. Spontaneous preterm births are those that arise from either spontaneous preterm labor or preterm premature rupture of membranes and account for two thirds of all preterm births. Indicated preterm births are those that occur because of a maternal or fetal indication for delivery. Indicated preterm births account for approximately one third of preterm births. Dr Bocking’s work focuses on spontaneous preterm births.

As mentioned in earlier text, spontaneous preterm labor accounts for a substantial fraction of all preterm deliveries. However, the clinical diagnosis of preterm labor is surprisingly fraught with inaccuracy. Let us take a common clinical example that is seen daily on a busy labor suite. Consider a woman who is a Gravida 1 Para 0 at 26 weeks’ gestation, with no risk factors for preterm birth. She presents with frequent uterine contractions, a reassuring fetal heart rate tracing, and a cervical exam of 1 cm/50%. Is this patient in preterm labor? Should we admit this patient? Should we use tocolytics? Should we give steroids? What should we tell the parents about the likelihood of delivery? Certainly, at 26 weeks we do not want to make a mistake and undertreat because if we are wrong and the baby is delivered soon, serious morbidity is likely.

In a clinical setting, accurate diagnosis of preterm labor is essential. As we can see from my hypothetical case, we need to make informed decisions about the use of tocolytics. All tocolytics can be associated with serious adverse events, and we want to limit their use to cases in which they are clearly needed. Likewise, we want to use antenatal steroids wisely. In days past, we would use serial courses of steroids in high-risk pregnancies, and now there is concern that multiple courses may increase rates of cerebral palsy. So we must use steroids only when the likelihood of preterm birth is high.
From a resource standpoint, we want to admit only patients who are at risk of preterm birth and avoid unnecessary hospitalizations and interventions. At Washington University, for example, we have a real crunch on beds, and we need to admit such patients carefully. Lastly, accurate diagnosis can inform future interventional clinical trials, enrolling only the highest-risk patients.

As was alluded to in earlier text, our current methods of diagnosing preterm labor are fraught with inaccurately, specifically a high false-positive rate. Traditional clinical methods include the assessment of uterine contractions and cervical change. This method is inaccurate, and we know this from the placebo-controlled trials of tocolytics. When clinical criteria are used, it has been shown that 50-80% of those randomized to placebo will deliver at term. Thus, the positive predictive value of clinical criteria is woefully low. Because of this inaccuracy, others have looked at more sophisticated measures.

Fetal fibronectin has been tested as a diagnostic test for preterm labor but was tested in a population with a low prevalence of preterm birth. Still, many believe that the high negative predictive value of a fetal fibronectin can help to avoid unnecessary hospitalizations in women with symptoms of preterm labor. Although this sounds good in theory, clinical trials have not supported that physiology to develop his decision trees. First, the use of biologically plausible biomarkers is wise. Clearly, clinical criteria and existing measures do not provide adequate discrimination, and we must consider new factors, such as maternal WBC or CRH, as additional factors. Second, Dr Bocking develops separate predictions based on gestational age at presentation. This is also very wise because it is largely believed, for example, that earlier preterm births are related to infection, whereas later preterm births are related to the activation of the hypothalamic-pituitary axis. This analytical approach is biologically driven, and Dr Bocking and his colleagues are to be commended for that. Next, Dr Bocking and colleagues chose a relevant clinical endpoint: delivery within 48 hours. This time frame is important for the administration of steroids. Next, as mentioned before, Dr Bocking uses a powerful statistical methodology to develop his decision trees. This is a methodology that we do not see much in obstetrics/gynecology but hope to in the future. Lastly, and perhaps most importantly, Dr Bocking validates his prior decision rules in a new population.

As we heard in Dr Bocking’s presentation, the results from his decision trees hold great promise, with favorable test characteristics. I would note several strengths of Dr Bocking’s study. First, the use of biologically plausible biomarkers is wise. Clearly, clinical criteria and existing measures do not provide adequate discrimination, and we must consider new factors, such as maternal WBC or CRH, as additional factors. Second, Dr Bocking develops separate predictions based on gestational age at presentation. This is also very wise because it is largely believed, for example, that earlier preterm births are related to infection, whereas later preterm births are related to the activation of the hypothalamic-pituitary axis. This analytical approach is biologically driven, and Dr Bocking and his colleagues are to be commended for that. Next, Dr Bocking and colleagues chose a relevant clinical endpoint: delivery within 48 hours. This time frame is important for the administration of steroids. Next, as mentioned before, Dr Bocking uses a powerful statistical methodology to develop his decision trees. This is a methodology that we do not see much in obstetrics/gynecology but hope to in the future. Lastly, and perhaps most importantly, Dr Bocking validates his prior decision rules in a new population.

Recursive partitioning is a predictive modeling technique: it maximizes the separation between groups of subjects, based on an outcome variable. And it does this by sequentially identifying a set of best possible splits to achieve that result. It is like stepwise logistic regression, but it has the added ability to look at specific levels of continuous and categorical variables to split on. The output is also different than stepwise logistic regression: instead of a model, it is a decision tree that was made from the best splits, and the final nodes are the product of the splits leading to that point.

As we heard in Dr Bocking’s presentation, the results from his decision trees hold great promise, with favorable test characteristics. I would note several strengths of Dr Bocking’s study. First, the use of biologically plausible biomarkers is wise. Clearly, clinical criteria and existing measures do not provide adequate discrimination, and we must consider new factors, such as maternal WBC or CRH, as additional factors. Second, Dr Bocking develops separate predictions based on gestational age at presentation. This is also very wise because it is largely believed, for example, that earlier preterm births are related to infection, whereas later preterm births are related to the activation of the hypothalamic-pituitary axis. This analytical approach is biologically driven, and Dr Bocking and his colleagues are to be commended for that. Next, Dr Bocking and colleagues chose a relevant clinical endpoint: delivery within 48 hours. This time frame is important for the administration of steroids. Next, as mentioned before, Dr Bocking uses a powerful statistical methodology to develop his decision trees. This is a methodology that we do not see much in obstetrics/gynecology but hope to in the future. Lastly, and perhaps most importantly, Dr Bocking validates his prior decision rules in a new population.

Oftentimes such rules work only in the data set in which they were created, so validation in some way is critical. With all of those strengths in mind, I do have some questions for Dr Bocking.

1. Please describe derivation population or define “threatened” preterm labor.
2. Please comment on other clinical factors that could be useful: history of preterm birth, ethnicity, etc. Could these improve the accuracy?
3. Is it possible that cointerventions (tocolytics, antibiotics) may distort test accuracy?
4. Is this population generalizable to other industrialized populations? Is there a role for external validation?
5. Can you provide 95% confidence intervals around the test characteristics?
6. Which decision trees can be used clinically today?
7. Comment on future research directions.

Dr Bocking (Closing). I thank Dr Macones for his supportive comments and important questions. I will address them as follows:

1. Please describe the derivation population (cervical dilation, contraction frequency) and define threatened preterm labor.

All women included in the Derivation population had presented to St Joseph’s Health Centre, London, Ontario, Canada, which is a tertiary perinatal center serving approximately 1 million people. They all had a clinical diagnosis of threatened preterm labor including the presence of regular uterine contractions, cervical dilation, and/or ruptured membranes. In 80% of cases, cervical dilation was less than 2 cm.

2. Please comment on other clinical factors that could be useful in prediction (ie, history of preterm delivery and ethnicity). Could these improve accuracy?

We included 15 variables in the initial analysis, which were chosen on the basis of the previous literature findings and/or...
biologic plausibility. The vast majority of women recruited to this study were Caucasian, and therefore, we can not assess ethnicity as a predictive factor for delivery within 48 hours. A history of prior preterm birth was not found to be significant for delivery within 48 hours.

3. Is it possible that cointerventions (tocolytics, antibiotics) may distort test accuracy?

It is unlikely that cointerventions such as tocolytics and/or antibiotics could have distorted our results because tocolytics were used infrequently in both groups and only in women who presented less than 32 weeks’ gestation. In the derivation cohort, magnesium sulphate was used. In the validation cohort, either ritodrine or indomethacin was used. In both groups, tocolytics were administered for only 48 hours.

4. Is this population generalizable to other industrialized populations? Is there a role for external validation?

Absolutely, there is a role for replicating these studies in other populations. The populations studied here were largely Caucasian women from middle socioeconomic status. The results of recursive partitioning are very dependent on the data obtained in the study population and therefore require replication prior to clinical application.

5. Can you provide 95% confidence intervals around test characteristics in the validation population?

We used both logistic regression and decision tree analysis to examine potential biomarkers in a complementary fashion for the derivation population. Because we did not perform logistic regression in the validation data, we can not provide 95% confidence intervals for the test characteristics in this group.

6. Which decision trees can be used clinically today? Under what circumstances?

Before these decision trees could be applied clinically, it will be important to replicate these studies in larger sample sizes and different populations. The results of this analysis would be best applied to symptomatic women in threatened preterm labor. Corticotrophin-releasing hormone (CRH) values should be normalized to finer gestational age categories from 32 to 36 weeks’ gestation because there is a large increase in CRH concentrations, even on a weekly basis at this gestational age. Recursive partitioning could be applied to a combination of the biomarkers assessed in this study with other clinically available diagnostic tools such as ultrasonic assessment of cervical length and cervicovaginal fetal fibronectin measurements. In addition, it would be of interest to look at the predictive value of decision trees on other clinically relevant endpoints such as delivery within 7 days of presentation, less than 32 weeks, and less than 37 weeks.

7. Comment on future research directions? Can computer based learning techniques be helpful?

There are a number of possible research directions for the future related to decision tree analysis for women in threatened preterm labor. Corticotrophin-releasing hormone (CRH) values should be normalized to finer gestational age categories from 32 to 36 weeks’ gestation because there is a large increase in CRH concentrations, even on a weekly basis at this gestational age. Recursive partitioning could be applied to a combination of the biomarkers assessed in this study with other clinically available diagnostic tools such as ultrasonic assessment of cervical length and cervicovaginal fetal fibronectin measurements. In addition, it would be of interest to look at the predictive value of decision trees on other clinically relevant endpoints such as delivery within 7 days of presentation, less than 32 weeks, and less than 37 weeks.