Predicting the Relation between Biopsychosocial Factors and Type of Childbirth using the Decision Tree Method: A Cohort Study

Saeideh Sadat Hajimirzaie1, MSc; Najmeh Tehranian2, PhD; Seyed Abbas Mousavi3, MD; Amin Golabpour4, PhD; Mehdi Mirzaii5, PhD; Afaneh Keramat6, PhD; Ahmad Khosravi7, PhD;

1Student Research Committee, School of Nursing and Midwifery, Shahroud University of Medical Sciences, Shahroud, Iran; 2Department of Reproductive Health and Midwifery, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran; 3Health Related Social and Behavioral Sciences Research Center, Shahroud University of Medical Sciences, Shahroud, Iran; 4Department of Basic Sciences, School of Medicine, Shahroud University of Medical Sciences, Shahroud, Iran; 5Department of Health Information Technology, Ophthalmic Epidemiology Research Center, Shahroud University of Medical Sciences, Shahroud, Iran; 6Department of Reproductive Health and Midwifery, Faculty of Medical Sciences, Shahroud University of Medical Sciences, Shahroud, Iran; 7Ophthalmic Epidemiology Research Center, Shahroud University of Medical Sciences, Shahroud, Iran

Correspondence: Ahmad Khosravi, PhD; Ophthalmic Epidemiology Research Center, Shahroud University of Medical Sciences and Health Services, Hafte Tir Sq., Shahroud, Iran; Tel: +98 23 32374350 Fax: +98 23 32395009 Email: a.khosravi@shmu.ac.ir

What's Known

- Previous studies showed that age, height, neonatal weight, high blood pressure, sugar, thyroid, toxemia, breech presentation, sleep disturbance, and multiple pregnancies are predictive factors for the childbirth method
- Further studies indicated 50 factors related to the childbirth method in four groups, pre-pregnancy, during the pregnancy, medical and social factors to predict the childbirth method

What's New

- The previous childbirth method, maternal body mass index at childbirth, maternal age, and serum estrogen were the most significant factors predicting the childbirth method
- The decision tree model was designed with high accuracy and sensitivity for predicting the childbirth method and by identifying the contributing factors and the rules derived from the model, health practitioners and policymakers can take preventive action.

Abstract

Background: With the growing rate of cesarean sections, rising morbidity and mortality thereafter is an important health issue. Predictive models can identify individuals with a higher probability of cesarean section, and help them make better decisions. This study aimed to investigate the biopsychosocial factors associated with the method of childbirth and designed a predictive model using the decision tree C4.5 algorithm.

Methods: In this cohort study, the sample included 170 pregnant women in the third trimester of pregnancy referring to Shahroud Health Care Centers (Semnan, Iran) from 2018 to 2019. Blood samples were taken from mothers to measure the estrogen hormone at baseline. Birth information was recorded at the follow-up time per 30-42 days postpartum. Chi-square, independent samples t test, and Mann-Whitney were used for comparisons between the two groups. Modeling was performed with the help of MATLAB software and C4.5 decision tree algorithm using input variables and target variable (childbirth method). The data were divided into training and testing datasets using the 70-30% method. In both stages, sensitivity, specificity, and accuracy were evaluated by the decision tree algorithm.

Results: Previous method of childbirth, maternal body mass index at childbirth, maternal age, and estrogen were the most significant factors predicting the childbirth method. The decision tree model’s sensitivity, specificity, and accuracy were 85.48%, 94.34%, and 83.87% in the training stage, and 83.33% in the testing stage, respectively.

Conclusion: The decision tree model was designed with high accuracy successfully predicted the method of childbirth. By recognizing the contributing factors, policymakers can take preventive action.

It should be noted that this article was published in preprint form on the website of research square (https://www.researchsquare.com/article/rs-34770/v1).


Keywords: Cesarean section ● Estrogens ● Biological factors ● Socioeconomic factors

Introduction

It is estimated that the world’s population will double over 60 years (two billion by 2050). In many countries, including Iran, due to an aging population, population control policies have become the incentive for childbearing. One reason for this decrease in...
Cesarean section (CS) will intensify the process of decrease in fertility rates by prolonging the interval among subsequent pregnancies and increasing the probability of secondary infertility. Moreover, placenta praevia, placenta accrete, and uterine rupture can be significantly increased in subsequent pregnancies after a primary CS.

Childbirth is a multidimensional process and a crucial experience in a mother's life, which is a unique experience while being universal. In certain instances, the childbirth method is determined based on mental health problems, ignorance, misconceptions, and attitudes, not based on medical indications. Fear of labor pain is a most common and serious problem among women. Psychological factors such as fear and stress can intensify labor pain, but it can be moderated by social support and self-efficacy. Moreover, higher maternal socioeconomic status and supplemental health insurance have shown to improve CS rate. Biological factors such as known hormones that contribute to labor progression (e.g., oxytocin and cortisol) are also linked to the psychological phenomena that may involve labor abnormality. Labor abnormality and increased CS are also associated with the estrogen hormone. Pregnancy is a hyperestrogenic state, and the placenta is a major source of estrogen secretion. Therefore, complex biological, psychological, and social factors influence the childbirth method, which can be clarified through a biopsychosocial model.

With the classification method, as a type of data mining, we can design a predictive model by identifying the different biopsychosocial variables. Classification is the process of finding a model that can identify unknown categories of other objects in order to identify the categories of data or data concepts. One of the common methods of classification is the decision tree. The decision tree is created by if-then rules for classification. Several algorithms are used in the decision tree construction, including ID3, C4.5, CHAID, and CART, the most important of which is the C4.5 algorithm.

Considering the increasing rate of CS in Iran, the effect of the childbirth method on different aspects of health, and the limited studies in Iran to predict the childbirth method, and since knowing the causes of CS can be a step toward reducing its rate, we decided to use the biopsychosocial model to investigate the associated factors and, by identifying these factors, design a model to predict the childbirth method using the C4.5 algorithm.

**Materials and Methods**

This cohort study was conducted on 170 pregnant mothers, that were referred to Shahroud Health Care Centers (Semnan, Iran) for pregnancy care during the third trimester of pregnancy from 2018 to 2019. The protocol of this study was reviewed and approved by the Institutional Review Board of Shahroud University of Medical Sciences (ethical code IR.SHMU.REC.1397.84). Explanations regarding study objectives and other necessary explanations were given to all participants, and written informed consents were taken for the publication of their clinical details. The inclusion criteria were as follows: Iranian citizenship, singleton pregnancy, having electronic records in the health system, no history of illness, and no onset of labor pain. A questionnaire, including demographic and socioeconomic variables, pregnancy history, and the WHO-5 well-being index, was completed.

**Measurements**

WHO-5 well-being index: The WHO-5 questionnaire consists of five questions about the participant’s feelings during the previous two weeks, with each item being scored based on a 6-point Likert scale of 0-5, which was initially developed to evaluate the quality of care for diabetic patients. The validity and reliability of the Persian version of this scale were assessed in the Iranian population. The validity of the WHO-5 questionnaire was assessed in a study on Iranian pregnant mothers up to eight weeks postpartum. Cronch’s alpha for WHO-5 items was 0.85, and a score of 50 or less with a good sensitivity of 84% and a specificity of 59% was used to identify the psychological symptoms. A 2015 review identified WHO-5 as a valid tool for depression screening. In our study, this scale was evaluated to have substantial reliability (that is, 0.81).

Socioeconomic index: The socioeconomic status was constructed using principal component analysis (PCA) according to the method described by Vyas and colleagues, which combines three main factors, including economic indicators (occupation, spouse occupation, homeownership status, a separate bedroom for couples, number of bedrooms, indoor bathroom, cooking area), asset-index (refrigerator, freezer, color TV, washing machine, dishwasher, microwave, vacuum cleaner, personal car, landline, mobile phone, computer or laptop, internet access), and social factors (education, spouse education, family members, family supplementary health.
insurance). At last, 15 variables were used to construct the socioeconomic status variables, which explained 17% of the total variance in this new variable. After calculating a variable called socioeconomic status, due to the twenty-fifth and seventy-fifth percentiles of these variables, the population was divided into high, medium, and low socioeconomic categories.

The height and weight of the mothers and their attendance in childbirth preparation classes were recorded. Then, 3 mL blood samples were taken from the antecubital vein in a non-fasting condition (from 9 a.m. to 11 a.m.) to measure the estrogen hormone. The blood samples were immediately transferred to the laboratory and centrifuged by the laboratory officer. Before analysis and after plasma separation, the samples were frozen at -80 °C. After ensuring that childbirth data were collected, serum estradiol E2 levels were measured by the enzyme-linked immunosorbent assay (ELISA) (Monobind kit, China).

Childbirth information was recorded at the time of referral to the health centers at 30-42 days postpartum. Participants’ blood samples were examined at the end of the second stage.

It should be noted that through the cooperation, mothers’ questions about the research or problems related to pregnancy and so on were answered by phone. Counseling was provided in case of any problems and referrals were given, if further counseling or treatment was needed.

Input Variables: Maternal age, family socioeconomic status, the previous method of childbirth, maternal mental health status during pregnancy, maternal body mass index (BMI) at childbirth, and maternal serum estradiol E2.

Output Variables: Method of childbirth, including VD and CS.

Model Design: At this point, the data were divided into training and testing datasets using the 70-30% method, meaning that 70% of the data were deemed as the training data, and 30% were considered as the testing dataset. The decision tree algorithm was implemented and formed based on the training data. The decision was then evaluated based on training and testing datasets. Finally, regarding the foregoing and considering the known risk factors of the CS, such as old age, obesity, history of CS, and failure to perform vaginal delivery after cesarean (VBAC), the rules were reviewed by the authors of the article, and the ones that were clinically valid and scientifically available were presented as the final rules.

Results

Nulliparous mothers accounted for 43.6% of the sample; 41.1% were primiparous and 15.4% multiparous. Only 27% of the mothers had attended childbirth preparation classes. The childbirth method of half of them (48.5%) was CS (35.6% non-emergency and 12.8% emergency). The causes of emergency CS were as follows: intrauterine growth restriction (IUGR), borderline amniotic fluid index (AFI), tachycardia, uterine malformations, and breech presentation (4.8% each), cephalopelvic disproportion (CPD) (9.5%), non-response to induction (19%), and meconium-stained amniotic fluid (MSAF) (47.6%). There were 34 cases of repeated non-emergency second CS. In the case of four subjects due to emigration from Shahroud and 10 subjects due to non-cooperation, the required information was obtained by telephone. Finally, 163 subjects were analyzed (figure 1). As seen in table 1, the mean±SD of variables such as age, previous method of childbirth, and maternal mental health status in pregnancy were significantly different between the two groups. Still, the socioeconomic status, serum estradiol level, gestational age at sampling time, and maternal BMI at childbirth were not significantly different.

Statistical Analysis

The data were analyzed using SPSS-23 (IBM, U.S.) and MATLAB 2017 (Mathworks, U.S.) software. Chi-square, Independent-Samples t test, and Mann-Whitney were used for the initial comparisons between the two groups. A P value<0.05 was considered as the significance level. Using input variables and determining the target variable, a C4.5 decision tree algorithm was developed.

Figure 1: The flow chart illustrates the process of recruitment and follow-up of a pregnancy cohort.
Bio-psychological factors affecting childbirth

### Model Construction

Training phase: In this phase, 70% of the 163 subjects were used in the training model. With the help of training data, the decision tree was created (table 2). However, the accuracy of these measures was to be tested in the testing phase.

In a decision tree, the most important information is the set of rules, consisting of several logical implications (If-then rule). Figure 2 shows the tree structure and presents the set of rules created by the C5 model.

Evaluation phase: A total of 48 samples were used for testing. The model output in the testing phase was compared with the actual results (table 2). The data were used to predict the decision tree rules. Some of the most important rules that were approved are as follows:

1. Mother's previous method of childbirth can predict the next method of childbirth with 95.65% probability.
2. A nulliparous mother with a BMI of less than 24.55 gives a VD with a 66.67% probability.
3. A nulliparous mother with ≥20.5 years of age, a BMI≥24.55 at the time of childbirth, and estradiol E2≥4702 pg/ml will have a VD with a probability of 88.89%.
4. A nulliparous mother with ≥20.5 years of age, a BMI≥24.55 at the time of childbirth, and estradiol E2 levels below 1097.1 pg/ml will have a CS with 66.67% probability.

### Discussion

Given the specific variables of a pregnant mother, CS probability is estimated based on the obtained rules. The previous method of

---

**Table 1: The comparison of demographic and socio-economic variables between two groups of childbirth methods**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total N=163</th>
<th>VD N=84</th>
<th>CS N=79</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr.) (Mean±SD)</td>
<td>29.08±5.76</td>
<td>27.9±5.89</td>
<td>30.33±5.38</td>
<td>0.007 *</td>
</tr>
<tr>
<td>BMI at Birth (kg/m²) (Mean±SD)</td>
<td>30.26±4.08</td>
<td>29.81±4.2</td>
<td>30.73±3.92</td>
<td>0.15 *</td>
</tr>
<tr>
<td>Estradiol E2 (pg/ml) (Mean±SD)</td>
<td>3407.02±1388.99</td>
<td>3481.01±1352.86</td>
<td>3328.35±1430.83</td>
<td>0.48 *</td>
</tr>
<tr>
<td>Gestational age (week) (Mean±SD)</td>
<td>31.81±3.32</td>
<td>31.72±3.32</td>
<td>31.9±3.34</td>
<td>0.7 *</td>
</tr>
<tr>
<td>Socioeconomic status index (%)</td>
<td>Low</td>
<td>41 (25.2)</td>
<td>21 (25.0)</td>
<td>20 (25.3)</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>82 (50.3)</td>
<td>45 (53.6)</td>
<td>37 (46.8)</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>40 (24.5)</td>
<td>18 (21.4)</td>
<td>22 (27.8)</td>
</tr>
<tr>
<td>Previous childbirth method (%)</td>
<td>None</td>
<td>71 (43.6)</td>
<td>43 (51.2)</td>
<td>28 (35.4)</td>
</tr>
<tr>
<td></td>
<td>VD</td>
<td>45 (27.6)</td>
<td>41 (48.8)</td>
<td>4 (5.1)</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>45 (27.6)</td>
<td>0</td>
<td>45 (57.0)</td>
</tr>
<tr>
<td></td>
<td>VD &amp; CS</td>
<td>2 (1.2)</td>
<td>0</td>
<td>2 (2.5)</td>
</tr>
<tr>
<td>Maternal mental health status (%)</td>
<td>≥50</td>
<td>26 (16.0)</td>
<td>9 (10.7)</td>
<td>17 (21.5)</td>
</tr>
<tr>
<td></td>
<td>&gt;50</td>
<td>137 (84.0)</td>
<td>75 (89.3)</td>
<td>62 (78.5)</td>
</tr>
</tbody>
</table>

*Independent-Samples t test; aMann-Whitney; bChi-square; VD: Vaginal delivery; CS: Cesarean section; BMI: Body mass index

---

**Table 2: The output of tree algorithm for training and test datasets**

<table>
<thead>
<tr>
<th>Confusion Matrix</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>PPV</th>
<th>NPV</th>
<th>Confusion Matrix</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>9</td>
<td>85.48%</td>
<td>94.34%</td>
<td>89.57%</td>
<td>94.64%</td>
<td>84.75%</td>
<td>14</td>
<td>3</td>
<td>82.35%</td>
<td>83.87%</td>
<td>83.33%</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>85.48%</td>
<td>94.34%</td>
<td>89.57%</td>
<td>94.64%</td>
<td>84.75%</td>
<td>5</td>
<td>26</td>
<td>82.35%</td>
<td>83.87%</td>
<td>83.33%</td>
</tr>
</tbody>
</table>

---

**Figure 2:** Flowchart demonstrates the proposed tree structure. VD: Vaginal delivery; CS: Cesarean section; BMI: Body mass index
childbirth, maternal BMI at childbirth, maternal age, and serum estrogen were the most significant variables predicting the type of childbirth, and the decision tree model predicted the method of childbirth with 89.57% accuracy in the training phase and 83.33% accuracy in the testing phase. In our study, there were no cases of VBAC; in other words, 100% of the CSs were performed without trying labor first. This was in line with results from many other studies, which concluded that CS in the first pregnancy leads to CS in later pregnancies, with a high probability.\textsuperscript{12, 22-25} Moreover, in line with our study, findings have also suggested that obesity is a contributing factor.\textsuperscript{26} Prior studies indicated that obesity led to a two to three-fold increase in the risk of emergency CS in primiparous and multiparous women.\textsuperscript{27} Furthermore, according to the present study and other researches, maternal age, especially in the first birth, is an important factor affecting CS.\textsuperscript{12, 28, 29} Estrogen hormone is another factor associated with labor abnormality and the CS thereafter. Estrogens induce myometrial changes, including increased production of prostaglandin (PG) E2 and F2α by the expression of the PG receptor enhancer, increased expression of the oxytocin receptor, an adrenergic agonist modulating membrane calcium channels, increased connexin synthesis, gap junction formation in the myometrium, and up-regulation of enzymes responsible for muscle contraction, such as light chain kinase myosin and calmodulin. All these changes result in synchronous uterine contractions that may facilitate VD.\textsuperscript{13}

In our study, maternal mental health in the third trimester of pregnancy and socioeconomic status were not predictive factors for the childbirth method. However, studies suggest that mental health problems before and during pregnancy, higher socioeconomic status, and possession of private health insurance are associated with higher CS rates.\textsuperscript{30-33}

Based on a previous study, the decision tree model was able to predict the childbirth method by an accuracy of 100% and 99.5 % in training and testing datasets. In this study, age, height, neonatal weight, high blood pressure, sugar, thyroid, toxemia, breech presentation, sleep disturbance, and multiple pregnancies were the predictive factors of the childbirth method.\textsuperscript{12} Another study indicated 50 factors related to the childbirth method in four groups, pre-pregnancy, during the pregnancy, medical, and social factors to predict the childbirth method. In that study, the factors most predictive of CS were medical factors. The decision tree model and artificial intelligence classified 80% and 82% of the childbirth results, respectively.\textsuperscript{24} Moreover, other models were also used to predict the method of childbirth. Factors affecting maternal VD in another study were the age of mother (20-25 years old), multiparity, unscarred uterus, rhesus positivity, expected baby weight of 2.5-3.5 kg, cephalic presentation, Bishop status>4, spontaneous onset of labor, clear liquor, and no fetal heart rate (FHR) abnormality. The composite score of ≤21 predicted VD with 80% sensitivity and 65% specificity. Still, a score above 21 did not always indicate CS.\textsuperscript{29}

This was a cohort study, and we had a sample loss of 8%. To solve this problem, we had to call the mothers on the phone to acquire information about the childbirth method, which reduced the information's accuracy. On the other hand, due to the low rate of delivery in emergency CS with labor, emergency CS without labor, and elective CS groups, we had to place everyone in one group (CS).

**Conclusion**

Data mining techniques can be used in the health care system by developing appropriate models to predict the likelihood of disease occurrence. Data collected from pregnant mothers, including estrogen hormone, were analyzed, and a high-precision model was designed using the C5 algorithm. To the best of our knowledge, this study is the first one to evaluate the estrogen hormone as a predictor of the childbirth method. In order to reduce the rate of CS in screening programs, it is recommended to implement this model with a larger sample size, after reaching the desired accuracy level. Health practitioners and policymakers can take preventive measures by identifying the contributing factors and the rules derived from this model.

**Acknowledgement**

Hereby, the authors of this article present their thanks and appreciation to Shahroud University of Medical Sciences. The present study was supported by Shahroud University of medical sciences as a PhD Thesis (grant No. 9740).

**Conflict of Interest:** None declared.

**References**


20 Vyas S, Kumarayake L. Constructing