Original Article

Risk Score for Prediction of Preterm Labor Inhibition Failure

Pavarin Teekapakvisit, MD[1], Pimpika Tansupswatdikul, MD[1]

[1] Department of Obstetrics and Gynecology, Chonburi Hospital, Chonburi, Thailand

Objective: To develop a risk scorecard for prediction of failure in preterm labor inhibition by selecting the most significant risk factors.

Materials and Methods: Retrospective cross-sectional analytic study collected 200 medical records of pregnant women between $24^{0/7}$ and $33^{6/7}$ weeks of gestation with the diagnosis of preterm labor who received tocolytic drugs at Chonburi Hospital. The 200 data were split into train and test datasets. Seventy percent (140) were allocated for training the model. The remaining 30% (60) were for testing the model's accuracy. The present study analyzed risk factors related to failure of preterm labor inhibition and selected those with most predictive power. Finally, a risk scorecard was developed using logistic regression.

Results: The significant risk factors were age, previous preterm labor, cervical dilatation, cervical effacement, duration, and interval of uterine contraction. Models were created and tested for accuracy of prediction of failure in preterm labor inhibition. The final model achieved 90% of AUC, 89% of sensitivity, and 78% of specificity in the test dataset. These four risk factors, which are history of previous preterm labor, cervical dilatation, duration, and interval of uterine contraction, were included in the final scorecard, and ranged from 0 to100. The score of 0 means high risk and 100 means low risk of failure in preterm labor inhibition.

Conclusion: The authors developed a scorecard that can accurately classify preterm labor patients into two groups, failed and successful. The model was tested and proved that it has good AUC, sensitivity, and specificity statistics.

Keyword: Inhibition of preterm labor, Risk score, Tocolytic drugs

Received 16 Oct 2018 | Revised 7 Jan 2019 | Accepted 8 Jan 2019

J Med Assoc Thai 2020;103(8):748-53
Website: http://www.jmatonline.com

Preterm birth is a common cause of neonatal morbidity and mortality. Preterm newborns often have both long- and short-term complications such as cerebral palsy, sensory deficit, and respiratory illness, comparing to term newborns. Each year, it is reported that preterm births are more than 10 percent of neonatal births, estimated at 15 million around the world. For Thailand, in 2010, there was an incidence for preterm birth at around 12%[1].

The National Institute of Health and the American College of Obstetrician and Gynecologists recommended a single course of corticosteroids for pregnant women between 24 and 34 weeks of gestation who are at risk of preterm delivery within seven days[2,3]. The use of tocolytic agents to suppress preterm uterine contraction can delay delivery for at least 48 hours, and this prolongation enable the complete administration of corticosteroids to obtain their maximum effect for gestational age of 24 to 34 weeks[3].

Nowadays, the authors do not have any equipment or methods to predict the failure rate of preterm labor inhibition with tocolytic drugs. The present study aimed to develop a method, which was a scorecard, to predict the failure rate of preterm labor inhibition. With the scorecard, the authors would be able to plan to appropriately refer patients with low success rate of preterm inhibition to tertiary centers. Tertiary centers are more appropriate for taking care of preterm newborns since they have obstetricians, pediatricians, and proper equipment. The development process and validation are described in the next sections.

Correspondence to:

Teekapakvisit P.

Department of Obstetrics and Gynecology, Chonburi Hospital, Muang, Chonburi 20000, Thailand.

Phone: +66-84-0668977

Email: kwang_koq@hotmail.com

How to cite this article:

Teekapakvisit P, Tansupswatdikul P. Risk Score for Prediction of Preterm Labor Inhibition Failure. J Med Assoc Thai 2020;103:748-53.

doi.org/10.35755/jmedassocthai.2020.08.9420

Materials and Methods

The present study was a retrospective cross-sectional analytics study and was approved by the Chonburi Hospital Ethics Committee for Human Research (no.38/2560). The authors collected medical records of singleton pregnant women between $24^{0/7}$ and $33^{6/7}$ weeks of gestation with the diagnosis of preterm labor who received tocolytic drugs and admitted at Chonburi Hospital, Thailand, between January 2016 and December 2017.

Preterm labor was diagnosed by ACOG criteria[3] (contraction of four times in 20 minutes or eight times in 60 minutes plus progressive changes of the cervix).

The authors excluded records with incomplete data and obstetric complications for contraindications to tocolysis such as oligohydramnios, hypertensive disorder in pregnancy, placenta previa, chorioamnionitis, fetal distress, and fetal congenital anomaly.

The present study collected three types of tocolytic drugs given to patients, which were 1) intravenous terbutaline, 2) oral nifedipine, and 3) intravenous terbutaline and oral nifedipine but not at the same time. Physicians applied a drug to a patient and if the uterine contraction did not stop, they decided whether to apply another drug. The use or sequences of the drug depended on the physician's opinion. The loading dosage of the intravenous terbutaline was 0.25 mg then 10 µg/minute and increase 5 µg/minute every 15 minutes to a maximum of 25 µg/minute. The loading dosage of nifedipine was 10mg every 15 minutes for three dosages then 10 to 20 mg every four to six hours. A single course of antenatal corticosteroids (four dosages of 6 mg intramuscular dexamethasone every 12 hours) was given to all patients.

The authors analyzed 15 risk factors for the present study. Those were also found in previous studies[4,5]. They were age, gestational age, numbers of gestation, previous preterm delivery, body mass index (BMI), cervical dilatation, cervical effacement, membrane rupture, interval of uterine contraction, duration of uterine contraction, white blood cell count, neutrophil, white blood cell in urine, rectovaginal swab culture, and tocolytic drug. The present study aimed to develop a scorecard to predict one target factor i.e., result of preterm labor inhibition (success or failure) and to prove the model could be applied to unknown dataset (test dataset).

Statistical analysis

The authors applied the Vapnik-Chervonenkis (VC) inequality[6] to determine the number of data points needed for the present study.

$$P[|E_{in} - E_{out}| > \varepsilon] \leq \delta$$

In this inequation, E_{in} is a percentage of error generated from the train dataset, E_{out} is a percentage of errors generated from the test dataset, ε is an approximation or desired error bound, and δ is the probability bound or 1 minus confidence interval (CI). The authors wanted to train a predictive model that could predict the train data nearly as good as the test data. The E_{in} and E_{out} should be closed for most of the time. For example, the inequation could be interpreted this way, when ε=0.1 and δ=0.05, it meant with a CI of 95%, the E_{in} and E_{out} were less 10% away from each other. In practice, the upper bound of δ could be simplified to $N^d e^{-N}$ where N was the number of data points needed to train a model, and d was VC dimension, which was numbers of risk factors plus one for logistic regressions. The $N^d e^{-N}$ was small when N was more than 10 times VC dimension. The authors had 15 factors. After selecting only the significant factors, the final model would contain less than 10 factors. With this assumption, the authors collected 200 medical records split into train (70%) and test (30%) datasets. Finally, the authors had 140 records to train the model and 60 to test the model. The authors used chi-square statistic to select the factors that are relevant to target factor (i.e., result of preterm labor inhibition). The authors used the GLM module[7] in R software[8] to train the logistic regression models. The authors tested the models using area under curve (AUC) of receiver operating characteristic (ROC), sensitivity, and specificity. The authors translated the developed model into scorecards that ranged from 0 to 100. The score of 100 means high probability of success.

Results

The characteristics of the train dataset separated into success and failure groups and p-values are shown in the Table 1.

The present study secondary outcome was to compare the differences in risk factors between the groups of success and failure in preterm inhibition. The authors observed that Apgar scores at 1 and 5 minutes after birth were indifferent between success and failure groups. However, the numbers of newborns with endotracheal tube intubation (ETT) and birth weight (BW) were significantly different across the groups. The secondary factors were tested by unequal variances t-test[9]. Table 2 shows the statistics of the tests. The success group received a

Table 1. The characteristics between success and failure groups

Factor	Success group (n=92, 66%) Mean±SD	Failure group (n=48, 34%) Mean±SD	p-value
Cervical dilatation (cm)	1.66±1.14	4.00±2.24	<0.001
Interval of uterine contraction (minute)	3.68±0.75	2.88±0.73	<0.001
Duration of uterine contraction (second)	42.93±7.99	49.06±4.57	<0.001
Cervical effacement (%)	64.18±22.93	77.81±17.71	<0.001
Age (year)	29.33±6.71	25.90±6.99	0.005
Previous preterm delivery; n (%)			0.031
Unknown (1st gestation)	38 (41)	20 (42)	
Yes	11 (12)	12 (25)	
No	43 (47)	16 (33)	
Tocolytic drug; n (%)			0.151
Nifedipine	25 (27)	13 (27)	
Terbutaline	48 (52)	31 (65)	
Both	19 (21)	4 (8)	
Gestational age (week)	30.76±2.38	30.18±2.43	0.176
Rectovaginal swab culture; n (%)			0.878
No growth	28 (30)	14 (29)	
Escherichia coli	26 (28)	10 (21)	
Gram post bacilli	18 (20)	14 (29)	
Other	20 (22)	10 (21)	
White blood cell	14,736.41±4,877.02	15,767.29±4,757.79	0.233
White blood cell in urine	5.29±11.18	6.04 ±11.61	0.362
Neutrophil (%)	81.06±8.35	81.78±9.56	0.660
Body mass index (kg/m^2)	25.37±5.17	24.83±5.25	0.574
Membrane rupture; n (%)			0.498
Yes	38 (41)	17 (35)	
No	54 (59)	31 (65)	
Numbers of gestation	1.93±1.05	1.96±1.11	0.901

SD=standard deviation

Table 2. Secondary outcome statistics

Factor	Success group	Failure group	p-value
Apgar at 1-minute	7.13	7.10	0.472
Apgar at 5-minute	8.39	8.28	0.348
ETT	20%	42%	0.003
BW (g)	1,817.37	1,489.79	<0.001

ETT=endotracheal tube intubation; BW=birth weight

complete course of dexamethasone, which possibly reduced the rates of ETT.

The authors completed the model development process by building models with all possible subsets of the six significant factors. The final model (Model-1) contained four risk factors, which were cervical dilatation, interval of uterine contraction, duration of uterine contraction, and previous preterm delivery. There were two more sample models. Model-2 had all significant six factors, which were cervical dilatation, interval of uterine contraction, duration of uterine contraction, cervical effacement, age, and previous preterm delivery. Model-3 had all the factors of Model-2 except for the previous preterm delivery factor. The statistics showing in Table 3 suggested that Model-1 was the best in term of AUC and sensitivity. In this case, the sensitivity was more important than the specificity because including false positive patient was safer than rejecting false negative patients.

Table 3. The models statistics in train dataset

Dataset	Model-1 train	Model-2 train	Model-3 train
AUC	88%	87%	88%
Sensitivity	83%	81%	78%
Specificity	75%	88%	87%

AUC=area under curve

Table 4. Scorecard table

Score	Odds	Probability of success
<10	0.05	5%
10 to 19	0.09	8%
20 to 29	0.16	14%
30 to 39	0.30	23%
40 to 49	0.55	35%
50 to 59	1	50%
60 to 69	2	67%
70 to 79	4	80%
80 to 89	8	89%
90 to 99	16	94%
100	32	97%

The authors converted the logistic regression model of the final into a scorecard as shown in Table 4. The authors designed this table so odds will be doubled when the score increases by 10 using the formula below[10]. The authors initially gave a base score of 50 for everyone.

$$\ln(\text{odds}) = \frac{\ln 2}{10}(\text{score} - 50)$$

The authors selected a threshold score of 60. If a patient's score was lower than 60, the authors predicted it as a failure inhibition, otherwise it was a success.

The score of 60 or above suggested that there was a 67% or more chance that it was a successful inhibition. A score for a patient consisted of a base score and subscores. The list of four subscores for the final model is shown in Table 5.

The following example, which was pulled from an actual data point, illustrates how to calculate a patient's risk score to predict preterm labor inhibition failure, as shown in Table 6. The example patient got a score of 47. When mapping the score to the designed score table, it shows that the patient had a low chance of 35% of successful inhibition. The actual result of the example was also a failure.

The authors also evaluated the developed models with the test dataset. The statistics for the test data are shown in Table 7. The final model (Model-1) still performed well with the test data. Figure 1 is the final model ROC for the test dataset.

For the final model, the statistics for the train and the test dataset were close, which meant the model could predict unknown data nearly as good as known data or even slightly better.

Table 5. Subscores for the factors

Factor	Examination	Subscore
Cervical dilatation	≤1	17
	2	4
	3	–4
	≥4	–24
Interval of uterine contraction	<3	–12
	3	–3
	≥4	10
Duration of uterine contraction	<40	14
	40 to 49	8
	≥50	–4
Previous preterm delivery	Unknown (1st gestation)	0
	No	2
	Yes	–4

Table 6. Score calculation example

Factor	Examination	Score
Cervical dilatation	2	4
Interval of uterine contraction	3	–3
Duration of uterine contraction	50	–4
Previous preterm delivery	Unknown	0
Base score		50
Total score		47

Table 7. The models statistics in test dataset

Dataset	Model-1 test	Model-2 test	Model-3 test
AUC	90%	85%	83%
Sensitivity	89%	68%	73%
Specificity	78%	90%	82%

AUC=area under curve

Discussion

The present study compared the means of factors between the two groups, the success inhibition group and the failure inhibition group. For continuous

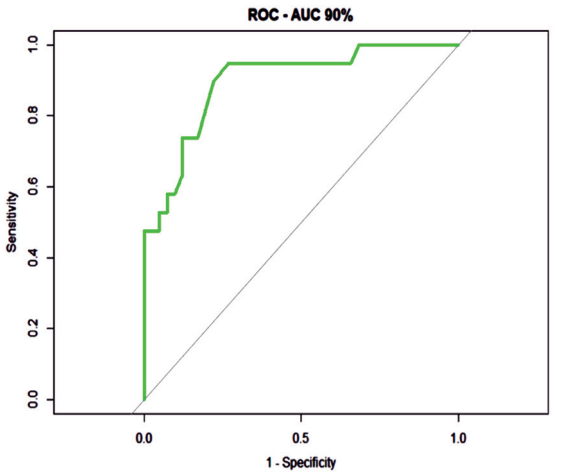

Figure 1. ROC for test dataset.

factors, the authors used unequal variances t-test. For categorical factors, the authors used the chi-square statistic. With p-values less than 0.05, it suggested that the means were different between the two groups.

There are four factors that are obviously different between the two groups according to their highly significant p-values. They are cervical dilatation, interval of uterine contraction, duration of uterine contraction, and cervical effacement. This similar finding was found in Kiatsuda et al[4] and Theplib and Phupong[5]. In addition, there are two more factors, which are age and previous preterm parity, considered different according to their p-values. Other factors are indifferent according to the test statistics.

The present study found four risk factors for the final model. They were cervical dilatation, interval of uterine contraction, duration of uterine contraction, and previous preterm delivery. Although, some of them were not the best in term of p-value, they collaboratively produced the best logistic regression model.

The designed scorecard ranges from 0 to 100, i.e., lowest to highest chance of success. The authors proposed a threshold score of 60, since it achieved good results in term of sensitivity (89%) and specificity (78%) in test data. Nevertheless, a threshold can be selected through discussions between tertiary and secondary centers. For example, a secondary center may not suitable for handling preterm births. A higher threshold score can be selected based on their agreement with a tertiary center to reduce unnecessary patient referrals to the tertiary center and effectively refer only cases with high possibility of failure in preterm inhibition.

The secondary outcome mentioned in the result section is also found in the previous study[11,12]. The failure group were born within 48 hours after received the drugs. On the other hand, the success group may have continued their pregnancies until term. This is probably be the rationale behind the BW test results. The present study was a retrospective cross-sectional analytics study. Limitations of this study were related to medical records collection. Each physician probably had different practice in taking care of preterm labor patients, such as first line of tocolytic drugs and dosages. In addition, there was no record related to side effect of tocolytic drugs such as hypotension, hyperglycemia, and tachycardia. Patients may have reacted differently to these side effects, which could lead to success or failure inhibition result.

Conclusion

The present study developed a predictive scorecard for prediction of failure in preterm labor inhibition. The developed scorecard was validated against samples hold out for validation, test dataset. The results confirmed that the scorecard developed can be used on incoming patients.

What is already known on this topic?

The use of tocolytic agents to suppress preterm uterine contraction can delay delivery for at least 48 hours, enough time to complete the administration of corticosteroids to obtain their maximum effect for preterm labor. However, there is no method to predict the failure rate of preterm labor inhibition.

What this study adds?

This study develops the predictive model to assist in decision making on referring a patient from primary or secondary center to tertiary centers if there was a high chance of failure inhibition.

With the scorecard, obstetricians can appropriately inform pediatricians if there was an imminent preterm labor. Therefore, they have time for preparations.

Acknowledgement

The authors would like to acknowledge Dr. Teera Siwadul, Head of Department of Obstetrics and Gynecology, Chonburi Hospital.

Conflicts of interest

The authors declare no conflict of interest.

References

1. Ruangdejnarong H, Tongsong T. Update in preterm

labor [Internet]. 2016 [cited 2017 Sep 1]. Available from: http://www.med.cmu.ac.th/dept/obgyn/2011/index.php?option=com_content&view=article&id=1265:update-in-preterm-labor&catid=45&Itemid=561.
2. National Institutes of Health Consensus Development Panel. Antenatal corticosteroids revisited: repeat courses - National Institutes of Health Consensus Development Conference Statement, August 17-18, 2000. Obstet Gynecol 2001;98:144-50.
3. American College of Obstetricians and Gynecologists' Committee on Practice Bulletins—Obstetrics. Practice Bulletin No. 171: Management of preterm labor. Obstet Gynecol 2016;128:e155-64.
4. Kiatsuda D, Thinkhamrop J, Prasertcharoensuk W. Success rate in preterm uterine contraction inhibition with tocolytic agents in a tertiary care center. Int J Womens Health 2016;8:663-7.
5. Theplib A, Phupong V. Success rate of terbutaline in inhibiting preterm labor for 48 h. J Matern Fetal Neonatal Med 2016;29:841-4.
6. Vapnik V. The nature of statistical learning theory. 2nd ed. New York: Springer; 2000.
7. R: Fitting generalized linear models [Internet]. 2017 [cited 2017 Sep 1]. Available at: http://stat.ethz.ch/R-manual/R-devel/library/stats/html/glm.html.
8. The R project for statistical computing [Internet]. 2017 [cited 2017 Sep 1]. Available from: https://www.r-project.org/.
9. Welch BL. The generalisation of student's problems when several different population variances are involved. Biometrika 1947;34:28-35.
10. Building credit scorecards using credit scoring for SAS® Enterprise MinerTM [Internet]. 2013 [cited 2017 Sep 1]. Available from: https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/building-credit-scorecards-using-credit-scoring-for-SAS-enterprise-miner-104182.pdf.
11. Garite TJ, Rumney PJ, Briggs GG, Harding JA, Nageotte MP, Towers CV, et al. A randomized, placebo-controlled trial of betamethasone for the prevention of respiratory distress syndrome at 24 to 28 weeks' gestation. Am J Obstet Gynecol 1992;166:646-51.
12. Mansouri M, Seyedolshohadaei F, Company F, Setare S, Mazhari S. Effect of antenatal Betamethasone on prevention of respiratory distress syndrome among neonates with gestational age of 35-36 weeks. J Gorgan Univ Med Sci [Internet]. 2010 [cited 2017 Sep 1]; 12:18-23. Available from: http://goums.ac.ir/journal/article-1-774-en.pdf.