Original Research

Validation of three models (Tolcher, Levine, and Burke) for predicting term cesarean section in Chinese population

Fangcan Sun[1,†], Minhong Shen[1,†], Bing Han[1,*], Youguo Chen[1], Fangfang Wu[1,2]

[1]The Department of Obstetrics and Gynecology, The First Affiliated Hospital of Soochow University, 215006 Suzhou, Jiangsu, China
[2]The Department of Obstetrics and Gynecology, Suzhou Ninth People's Hospital, 215200 Suzhou, Jiangsu, China
*Correspondence: hanbing@suda.edu.cn (Bing Han)
†These authors contributed equally.

Abstract

Background: Some models predicting cesarean section (CS) have been proposed, with Tolcher, Levine, and Burke model well acknowledged. Tolcher model targets nulliparous women with term labor induction; Levine model targets women with term labor induction with intact membranes and an unfavorable cervix. Burke model targets term nulliparous woman with an uncomplicated pregnancy. Our objective was to assess the predictive performance of these three models, and to disclose the variables which may predict the risk of CS in Chinese population. **Methods**: A retrospective study was conducted on women with singleton, term, cephalic pregnancies at a tertiary academic center (2011–2017). A predicted probability for CS was calculated for women in the dataset by the algorithm of each model. The performance of the model was evaluated for discrimination. Univariate analysis was used to screen out the factors that may increase the risk of CS. **Results**: The three models predicted CS as following (expressed by an area under the receiver operating characteristic curve [AUC ROC]) (in the population defined/employed by each model): Tolcher model with AUC ROC of 0.659; Levine model with 0.697; and Burke model as 0.623. Different interventional measures or characteristics of labor were also evaluated; the nulliparous and multiparous were analyzed separately. Still, most of the results were unsatisfactory (AUC ROC <0.7). Univariate analyses on the clinical parameters that may affect the incidence of CS were performed. The followings affected the incidence/probability of CS: maternal age, height, body mass index (BMI), weight gain during pregnancy, gestational age, mode of labor induction, meconium-stained amniotic fluid, presence of complications, neonatal weight/gender. **Conclusion**: These three models may not be suitable for predicting CS for Chinese population. Some maternal and fetal characteristics increased the risk of CS, which should be taken into account in creating some appropriate models for predicting CS in Chinese population.

Keywords: Cesarean section; Induction of labor; Nomogram; Prediction algorithm; Prediction tool

1. Introduction

Cesarean section (CS) is an important surgery within the obstetric domain. It has become an effective means to solve dystocia and some obstetric complications, and to save maternal and perinatal lives. However, it is not a normal mode of delivery. Apart from the risks of the surgical procedure itself, it could also cause short and long term complications [1,2]. In addition to have a profound impact on childbirth experiences of women, trial of labor has clinical and social implications due to their unpredictable duration and chance of success [3]. On the other hand, the childbirth experience among nulliparous women may affect her next pregnancy, especially with the introduction of China's universal two-child and three-child policies. Therefore, in order to improve their outcomes, providing the safe, appropriate and personalized delivery mode according to the view of mothers and their babies is vital.

Currently, several published models, which were based on maternal and fetal variables, were available to predict the probability of unplanned CS in women at term [4–8]. A study by Tolcher *et al*. [4] used a nomogram to predict the probability of CS. Tolcher model was developed using a retrospective cohort of 785 nulliparous undergoing induction of labor at term at Mayo Clinic Rochester and had a good discriminatory ability with a bias-corrected c-index of 0.709 (95% confidence interval [CI] 0.671–0.750). Levine *et al*. [5] developed and validated a predictive model for women undergoing an induction of labor at term with intact membranes and an unfavorable cervix (Bishop score ≤ 6 and dilation ≤ 2 cm). The Levine model was developed as part of a secondary analysis of a large randomized trial (n = 491) and was validated by an observational cohort (n = 362). A nomogram was created from the model with a c-index of 0.77 in the development cohort and 0.73 in the validation cohort. Burke *et al*. [6] used a nomogram to develop a predictive model in nulliparous woman from 39^{+0} to 40^{+6} weeks' gestation with an uncomplicated pregnancy. Burke model was developed in a prospective, multicenter, blinded observational study that recruited 2336 nulliparous women, and had an excellent calibration and discriminative ability with a misclassification rate of 0.21.

However, these prediction models' population is majority white. It is not clear whether these models are applicable to other ethnicity. To date, there is no method to accurately stratify pregnant women according to their risk profile for CS. The benefits of using these models should be demonstrated before routine introduction into clinical practice. Thus, the aim of study was to assess the predictive performance of three foreign models for Chinese population, and to discover the variables which may be useful to predict the risk of CS in this population.

2. Materials and methods

A retrospective study was conducted on women who fulfilled the following inclusion and exclusion criteria at a tertiary care academic center in the department of obstetrics and gynecology of the First Affiliated Hospital of Soochow University from January 1st, 2011 to August 31st, 2017. The protocol for this study was approved by the Institutional Review Board at this center (2018019).

Women with singleton, term (37 0/7 weeks of gestation or greater) and cephalic pregnancies were eligible for inclusion. Women were excluded if they had any one of the following: severe complications during pregnancy (cardiac failure, severe liver and kidney diseases, severe preeclampsia complicated with organ dysfunction), a scared uterus (prior CS delivery or myomectomy), complete or partial placenta previa or vasa previa, prolapse or presentation of umbilical cord, other contraindications to vaginal delivery. Women with the CS on maternal request were also excluded.

The outcome of interest was defined as CS. Data on maternal characteristics and perinatal parameters were collected from the institution's obstetrics database, which was obtained by the patient's medical record review. The following variables were recorded (Table 1): maternal parity, age, height, weight, baseline body mass index (BMI) (BMI was calculated as weight/height^2), weight change during pregnancy, gestational age at delivery (all participants had a confirmed estimated date of delivery by either first-trimester ultrasound or second-trimester ultrasound that correlated with their menstrual dates), pregnancy complications (hypertensive disorders of pregnancy, diabetes mellitus, intrahepatic cholestasis of pregnancy, polyhydramnios, oligohydramnios, uterine myoma etc.), premature rupture of membranes, epidural analgesia, meconium-stained amniotic fluid, induction methods (oxytocin, amniotomy, prostaglandin E2, disposable cervical dilator balloon), cervical dilation at induction, mode of delivery, major indications for CS, fetal ultrasound features (biparietal diameter, head circumference, abdominal circumference, and femur length which were measured within 1 week prior to delivery), newborn sex, and newborn birth weight.

We searched the literature for models that predicted the likelihood of CS among women with singleton, term, and cephalic pregnancies between January 1st, 2011 to August 31st, 2017, the same period as our study period. Considering the limitations of retrospective data collection, three studies [4–6] with relatively larger sample sizes were selected for external validation. These original articles were published in first-class obstetrics and gynecology journals and worthy. The detailed parameters of three models (Tolcher, Levine and Burke models) were presented in **Supplementary Table 1**.

A predicted probability for CS was calculated for woman in the dataset by means of published each model algorithm. The algorithms of the three models all used nomograms to calculate the probability of CS. For a given woman, each characteristic was aligned with the corresponding number of scores on the scores axis in the nomogram, and a total summated score was derived. The sum of all scores lined with predicted probability of CS in the nomogram.

Considering the induction/augmentation methods of labor may affect the delivery mode, we used the following methods for grouping based on the experience of the study institution. Firstly, women were divided into the Non-intervention group and the Intervention group according to whether they received intervention measures. The Non-intervention group was defined as women who entered labor naturally and the labor process did not be intervened. Then, women of the Intervention group were divided into the Augmentation group and Induction group according to whether the initial cervix was dilated by 6 cm. Women who received intervention measures such as amniotomy and/or oxytocin when their initial cervical dilation was greater than or equal to 6 cm were defined as the Augmentation group. Women who received intervention measures when their initial cervical dilation was less than 6 cm were defined as the Induction group. The Induction group was then divided into four subgroups according to different induction methods. (1) Oxytocin Induction group: Women only received oxytocin induction. (2) Amniotomy group: Women received artificial rupture of membranes with or without oxytocin induction. (3) Prostaglandin E2 group: Women received Prostaglandin E2 (Propess) induction. (4) Disposable Cervical Dilator Balloon group: Women received disposable cervical dilator balloon induction.

For continuous variables, the tests of normality was performed first. The Student's *t*-test was used to compare the continuous variables with a normal distribution. The chi-square test, Fisher exact test and Wilcoxon rank-sum test were used, as appropriate, for the categorical variables. Standard descriptive statistics (mean ± standard deviations) were used to summarize continuous variables. Percentage and frequency were used for categorical variables. All *p* values were two-tailed, and a significance level of 5% was used. The area under the receiver operating characteristic curve (AUC ROC) was calculated to assess the discrimination power of each model. AUC ROC was interpreted using following categories: non-informative (AUC ROC = 0.5),

Table 1. Characteristics of maternal and neonatal by mode of delivery.

Variable	Cesarean section (n = 725, among nulliparous)	Vaginal delivery (n = 6636, among nulliparous)	p*	Cesarean section (n = 34, among multiparous)	Vaginal delivery (n = 2379, among multiparous)	p*
Maternal age (y)	27.58 ± 3.16	26.90 ± 3.06	<0.001	32.44 ± 5.36	30.57 ± 3.90	0.006
Baseline height (cm)	159.64 ± 4.39	161.69 ± 4.62	<0.001	159.24 ± 4.40	161.10 ± 4.53	0.017
Baseline BMI (kg/m^2)	21.64 ± 2.74	20.88 ± 2.61	<0.001	23.36 ± 3.64	21.67 ± 2.71	<0.001
Weight change during pregnancy (kg)	14.59 ± 4.24	14.05 ± 4.27	0.001	13.38 ± 4.30	12.86 ± 4.27	0.488
BMI at delivery (kg/m^2)	27.37 ± 3.14	26.26 ± 2.88	<0.001	28.63 ± 3.55	26.63 ± 2.87	<0.001
Gestational age at delivery (w)	40.18 ± 1.00	39.74 ± 1.00	<0.001	40.00 ± 1.00	39.62 ± 1.00	0.015
HDP	36 (4.97%)	149 (2.25%)	<0.001	1 (2.94%)	41 (1.72%)	1.000
Diabetes mellitus			0.001			0.207
None	665 (91.72%)	6275 (94.56%)		29 (85.29%)	2184 (91.80%)	
Pre-existing	2 (0.28%)	2 (0.03%)		0 (0.00%)	1 (0.04%)	
GDM	58 (8.00%)	359 (5.41%)		5 (14.71%)	194 (8.15%)	
ICP	5 (0.69%)	23 (0.35%)	0.190	0 (0.00%)	12 (0.50%)	1.000
Amniotic fluid volume			<0.001			0.015
Normal	671 (92.55%)	6512 (98.13%)		31 (91.18%)	2338 (98.27%)	
Polyhydramnios	24 (3.31%)	55 (0.83%)		1 (2.94%)	23 (0.97%)	
Oligohydramnios	30 (4.14%)	69 (1.04%)		2 (5.88%)	18 (0.76%)	
Uterine myoma	6 (0.83%)	25 (0.38%)	0.139	0 (0.00%)	10 (0.42%)	1.000
PROM	175 (24.14%)	1395 (21.02%)	0.052	6 (17.65%)	379 (15.93%)	0.786
Intervention/augmentation methods			<0.001			<0.001
Non-intervention group plus Augmentation group	134 (18.48%)	2811 (42.36%)		7 (20.59%)	1242 (52.21%)	
Oxytocin Induction group	173 (23.86%)	1103 (16.62%)		6 (17.65%)	197 (8.28%)	
Amniotomy group	345 (47.59%)	2438 (36.74%)		18 (52.94%)	889 (37.37%)	
Disposable Cervical Dilator Balloon group	30 (4.14%)	99 (1.49%)		3 (8.82%)	42 (1.77%)	
Prostaglandin E2 group	43 (5.93%)	185 (2.79%)		0 (0.00%)	9 (0.38%)	
Epidural analgesia	29 (4.00%)	179 (2.70%)	0.044	0 (0.00%)	11 (0.46%)	1.000
Meconium-stained amniotic fluid			<0.001			<0.001
0	547 (75.45%)	6109 (92.06%)		24 (70.59%)	2162 (90.88%)	
I	42 (5.79%)	246 (3.71%)		1 (2.94%)	94 (3.95%)	
II	38 (5.24%)	163 (2.46%)		4 (11.76%)	61 (2.56%)	
III/ bloody	98 (13.52%)	118 (1.78%)		5 (14.71%)	62 (2.61%)	
Male infant	410 (56.55%)	3247 (48.93%)	<0.001	24 (70.59%)	1217 (51.16%)	0.024
Infant birth weight	3527.63 ± 401.10	3341.07 ± 374.99	<0.001	3754.41 ± 336.68	3452.55 ± 392.86	<0.001

BMI, body mass index; PROM, premature rupture of membranes; HDP, hypertensive disorders of pregnancy; ICP, intrahepatic cholestasis of pregnancy; GDM, gestational diabetes mellitus.

Data are mean $\pm$ standard deviation, n/(N %).

* Two-sided p based on the χ^2 or Fisher's exact or Wilcoxon rank-sum test for categorical variables, and the t test for continuous variables.

poor accuracy ($0.5 <$ AUC ROC < 0.7), moderate accuracy ($0.7 <$ AUC ROC < 0.9), high accuracy ($0.9 <$ AUC ROC < 1); and perfect accuracy (AUC ROC $= 1$) [9]. The cut-off point according to the Youden index (sensitivity + specificity-1) which reflected the best accuracy was selected. All analyses were performed with SPSS (IBM SPSS Statistics 24 for Windows, IBM Corp., Chicago, IL, USA) and GraphPad Prism software (Version 7.0 for Windows, San Diego, CA, USA).

3. Results

3.1 Cohort composition and demographics

During the study period, a total of 18,228 deliveries were delivered at the hospital. After exclusion of ineligible patients, 7361 nulliparous and 2413 multiparous women were enrolled. 725 nulliparous and 34 multiparous women underwent CS delivery. Non-reassuring fetal heart rate status was the most common indication for CS (nulliparous: 354/725 = 48.83%; multiparas: 18/34 = 54.94%) followed by arrest of active phase (nulliparous: 153/725 = 21.10%; multiparas: 8/34 = 23.53%).

The chi-square test was performed for maternal parity, and the result significant difference between the vaginal and cesarean group ($p < 0.001$). The CS rate of nulliparous and multiparous was 9.85% and 1.41%, respectively. Therefore, in this study, nulliparous and multiparous were analyzed separately.

3.2 Univariate analysis of the characteristics of the study population

Maternal and neonatal characteristics were compared between women who underwent CS and women who delivered vaginally (Table 1).

3.2.1 Results in nulliparous

There were 7361 cases of nulliparous, including 6636 (90.15%) vaginal deliveries and 725 (9.85%) CS deliveries. In univariable analysis, women who underwent CS delivery were more likely to be older, shorter stature, older gestational age, greater baseline and delivery BMI, more gestational weight gain. Women who underwent CS delivery also had more pregnancy-associated complications and meconium-stained amniotic fluid, had a higher rate of delivering with labor epidural analgesia, and had heavier newborn birth weight than those in the vaginal delivery group ($p < 0.05$). When the cut-off value was age >26 years old, height ≤ 160 cm, gestational age at delivery >40 weeks, pre-pregnancy BMI >21.3 kg/m^2, pregnancy weight gain >13 kg, and neonatal weight >3490 g, the risk of cesarean was significantly increased. The composition ratio of intervention methods were statistically significant between the two groups ($p < 0.05$). Although the rate of premature rupture of membranes was lower in the vaginal delivery group, there were not statistically significant ($p > 0.05$).

3.2.2 Results in multiparas

There were 2413 multiparas including 2379 (95.59%) vaginal deliveries and 34 (1.41%) CS deliveries. The maternal age, gestational age at delivery, BMI, rate of meconium-stained amniotic fluid, rate of male fetuses, and neonatal weight of the CS group were all higher than those of the vaginal delivery group ($p < 0.05$). The composition ratio of intervention methods were also statistically significant ($p < 0.05$). Compared with the vaginal delivery group, the CS group had more weight gain during pregnancy and higher incidence of premature rupture of membranes, but the differences between the two groups were not statistically significant ($p > 0.05$). The risk of CS was increased when the age was more than 33 years old, the height was less than 162 cm, the gestational age at delivery >40 weeks, pre-pregnancy BMI >22.4 kg/m^2, and newborn weight >3590 g. There were not significant difference between the two groups in the rate of delivering with labor epidural analgesia ($p > 0.05$).

3.3 Verify the models

Supplementary Fig. 1 describes the study population profile. As presented in Table 2 and Fig. 1, these results were performed to assess the performance of the model in various subsets of the study.

3.3.1 Verify the Tolcher model among nulliparous and multiparas

The Tolcher model was established to evaluate the probability of CS after induction of labor among nulliparous. In this study, there were a total of 7361 cases of nulliparous. Eighteen of these women had predictors exceed the range of the nomogram. In the end, 5630 nulliparous women who experienced labor intervention were participated externally validation. The calculated AUC ROC among nulliparous after induction of labor was 0.659 (95% CI 0.635–0.682), and the probability cut-off value of 0.286 (Fig. 1A-orange line). Then, different intervention measures of labor were evaluated separately. The AUC ROC of the Augmentation group, Oxytocin Induction group, Amniotomy group and Prostaglandin E2 group were 0.677, 0.635, 0.669 and 0.637, respectively (Fig. 1A).

Multiparas in the study were also evaluated. Among these women, the AUC ROC of the Induction group, Augmentation group, Oxytocin Induction group and Amniotomy group were 0.726, 0.894, 0.821 and 0.674, respectively (Fig. 1B). While the difference in the prediction probabilities of the Disposable Cervical Dilator Balloon group and Prostaglandin E2 group were not statistically significant.

3.3.2 Verify the Levine model among nulliparous and multiparas

The Levine model was developed to assess the probability of CS for women (nulliparous plus multiparous) un-

**Table 2. Performance of each model among nulliparous and multiparous women.**

	VD	CS	AUC ROC	95% CI	Cut-off
Tolcher					
Original research	554	231	0.709	0.671–0.750	
Validation					
Among nulliparous					
Augmentation group plus Induction group	5007	623	0.702	0.681–0.723	22.9%
Augmentation group	1194	34	0.677	0.587–0.767	11.8%
Induction group	3813	589	0.659	0.635–0.682	28.6%
(1) Oxytocin Induction group	1099	172	0.635	0.588–0.682	29.1%
(2) Amniotomy group	2433	344	0.669	0.639–0.699	24.3%
(3) Disposable Cervical Dilator Balloon group	99	30	0.563	0.449–0.677	
(4) Prostaglandin E2 group	182	43	0.637	0.540–0.735	48.5%
Among multiparous					
Augmentation group plus Induction group	1487	30	0.749	0.656–0.842	24.8%
Augmentation group	354	3	0.894	0.770–1.018	13.3%
Induction group	1133	27	0.726	0.617–0.834	38.1%
(1) Oxytocin Induction group	197	6	0.821	0.666–0.975	45.5%
(2) Amniotomy group	885	18	0.674	0.535–0.813	39.5%
(3) Disposable Cervical Dilator Balloon group	42	3	0.825	0.677–0.974	
(4) Prostaglandin E2 group	9	0			
Levine					
Original research	355	136	0.77		
Validation					
Among nulliparous plus multiparous					
Augmentation group plus Induction group	4074	477	0.726	0.704–0.748	25.4%
Only woman with intact membranes and an unfavorable cervix	956	175	0.697	0.656–0.738	29.0%
Augmentation group	906	26	0.672	0.567–0.777	23.6%
Oxytocin Induction group	843	134	0.691	0.647–0.736	25.4%
Amniotomy group	2097	257	0.737	0.708–0.766	25.4%
Disposable Cervical Dilator Balloon group	106	25	0.666	0.549–0.785	30.8%
Prostaglandin E2 group	122	35	0.647	0.540–0.755	60.8%
Among nulliparous					
Augmentation group plus Induction group	2982	452	0.664	0.639–0.688	30.8%
Augmentation group	648	23	0.635	0.515–0.755	33.5%
Oxytocin Induction group	692	129	0.648	0.598–0.699	29.9%
Amniotomy group	1454	243	0.665	0.630–0.700	31.7%
Disposable Cervical Dilator Balloon group	73	22	0.630	0.500–0.759	
Prostaglandin E2 group	115	35	0.626	0.513–0.738	61.7%
Among multiparous					
Augmentation group plus Induction group	1092	25	0.638	0.524–0.751	7.6%
Augmentation group	258	3	0.577	0.295–0.858	
Oxytocin Induction group	151	5	0.694	0.435–0.953	8.2%
Amniotomy group	643	14	0.645	0.483–0.807	5.2%
Disposable Cervical Dilator Balloon group	33	3	0.626	0.392–0.860	
Prostaglandin E2 group	7	0			
Burke					
Original research	1845	491	0.69		
Validation					
Among nulliparous puls multiparous	252	22	0.619	0.502–0.737	23.0%
Among nulliparous	180	19	0.623	0.500–0.746	23.0%

VD, vaginal delivery; CS, cesarean section; AUC ROC, area under the receiver operating characteristic curve; CI, confidence interval.

dergoing an induction of labor with intact membranes and an unfavorable cervix. In this study, there were 3575 cases of baseline BMI less than 20 kg/m^2. These variables exceed the range of the nomogram. We then excluded women who had not experienced labor intervention, and finally, 4551 cases were externally validated with Levine model. Among 1131 women after induction of labor with intact membranes and an unfavorable cervix, the AUC ROC was 0.697 (95%

CI 0.656–0.738), and the probability cut-off value of 0.290 (Fig. 1C-orange line). Similarly, according to the methods of intervention, the results showed that the AUC ROC was 0.672 in Augmentation group, 0.691 in Oxytocin Induction, 0.737 in Amniotomy group, 0.666 in Disposable Cervical Dilator Balloon group, and 0.647 in Prostaglandin E2 group (Fig. 1C). Also, nulliparous and multiparous were also separately analyzed, and the statistically significant re-
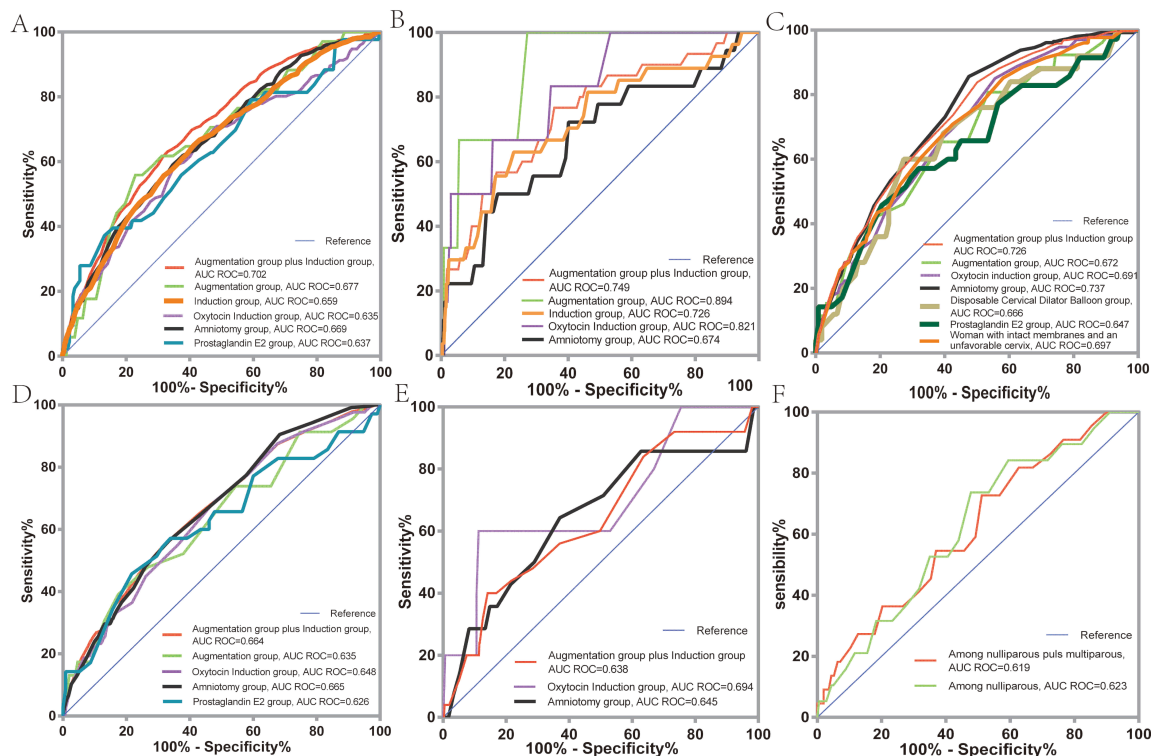
Fig. 1. Receiver operating characteristic (ROC) curves of each model for predicting cesarean section. (A) ROC curves of Tolcher model among nulliparous. (B) ROC curves of Tolcher model among multiparous. (C) ROC curves of Levine model among nulliparous plus multiparous. (D) ROC curves of Levine model among nulliparous. (E) ROC curves of Levine model among multiparous. (F) ROC curves of Burke model.

sults were presented in Table 2 and Fig. 1D–E. The discrimination of these groups were limited ($0.5 <$ AUC ROC < 0.7).

3.3.3 Verify the Burke model among nulliparous and multiparas

The Burke model was developed to calculate the probability of CS in nulliparous woman from 39^{+0} to 40^{+6} weeks' gestation with an uncomplicated pregnancy. In this study, there were 4813 women whose predictors exceed the range of the nomogram. And 4498 cases lacked the fetal ultrasound findings in the week before delivery. Then we removed the women with complications during pregnancy, finally, 274 cases (199 cases were nulliparous and 75 were multiparous) were validated. The AUC ROC was 0.623 (95% CI 0.500–0.746) among nulliparous. When multiparous women were added, the AUC ROC was 0.619 (Fig. 1F).

4. Discussion

With the change of China's one-child policy, an increasing number of researchers and pregnant women are commonly and widely concerned about the choice of delivery mode. The ability to predict the risk of CS for women with singleton, term and cephalic pregnancies would be highly beneficial to guide their management in labor. We searched the literature for models that predicted the probability of CS delivery among women with singleton, term, and cephalic pregnancies during the same period as our study. Three models with relatively high quality and large sample size were included. But, the effectiveness of a new predictive model must be evaluated before it put into clinical practice. Thus, we externally validated these models (Tolcher, Levine, and Burke models) in an existing cohort of Chinese women. At the same time, nulliparous and multiparous were evaluated separately. Different methods of induction were verified as well.

Unfortunately, under these specific conditions, the predictive abilities of three models were mostly poor (AUC ROC <0.70). This may be due to some limitations. The demographic makeup of our study population was different from those of the original research. Chinese women are relatively thin and have a low BMI. Then, pregnant women with complications were less likely to choose vaginal trial delivery. Moreover, different induction methods in different institutions may result in varying rates of CS. Therefore, the three tools for predicting CS may not be suitable for Chinese population in this tertiary care academic center. What's more, China has a vast territory and a large population, and the economy, medical level and life circumstances of these regions are greatly different, with a huge geographic variation in the cesarean rate [10]. So it's not

clear whether or not these models can be applied to settings in other parts of China.

We also preliminarily explored some variables that might be useful in predicting the risk of CS. In our study, maternal age, height, BMI, weight gain during pregnancy, gestational age, mode of labor induction, meconium-stained amniotic fluid, presence of complications, neonatal birth weight, and neonatal gender affected the delivery mode. Previous studies have shown that these factors at least partly determine the incidence of CS [11–13], and the gestational age of induced labor also affects it [14].

As far as we know, this is the first study to simultaneously concluding validations of above these three models for predicting CS delivery at term in Chinese population. These models were validated not only in a standard population, but also in multiparas. Unfortunately, the results suggest that these models are not suitable for the target population. The results are negative, but they show the necessity of validating existing models in different settings and populations before widespread implementation in clinical practice. The results also indicate that models need to be established locally and ethnically. Most prediction models may be only applicable to the specific populations which the original models normally target on. Additionally, further investigation of model validity and impact is important and should be undertaken.

Some limitations in this study should also be acknowledged. Firstly, the sample size was small, and this study had a retrospective design. Because of limitations of the quality of case records, some data is inevitably missing and inherently biased. Second, the cervical ripening was assessed by only cervical dilation during induction. We found in the literature review that Levine *et al*. [7] optimized their previous prediction model. The following five variables, which were modified Bishop score, gestation age $\geq$40 weeks, nulliparity, BMI at delivery, and height, were significantly associated with CS delivery in multivariable modeling. They found that the nomogram had AUC ROC in the derivation cohort of 0.79 and in the external validation cohort of 0.73. Meanwhile, they also had a user-friendly website for calculating. We did not verify this model, because the Bishop score were not completely collected. Dr D'Souza and colleagues [15] attempted to externally validate this model in the patient population at Mount Sinai Hospital. Yet the AUC ROC was 0.61 (95% CI 0.53–0.68) and it showed the performance was modest. Third, we excluded some patients who did not meet the criteria of the nomogram, but this subset of patients may potentially have an increased risk of CS. For example, when the sonographic head circumference of fetuses was greater than 360 mm, the risk of CS increased [16].

At present, many prediction models were established in many medical fields, but few are maturely implemented in clinical care. The performance, impact, and usefulness of prediction models need to be supported prior to prac-

tice [17]. Three models were externally validated in our center, but the results were not satisfactory. Given the diversity of geography, economy, medical level and environment across China, it is unclear whether these models can be applied to medical centers in other parts of the country. With the development of data technology, it is expected to conduct a large sample, multicenter, prospective study in China. And prediction models should be established accurately, conveniently and locally. It is worth noting that these models should not be used in isolation, but should be combined with the actual conditions of patients.

5. Conclusions

The Tolcher, Levine and Burke models for predicting CS may not be suitable for Chinese population in this hospital, but the applicability of these models in this population needs to be further explored with a larger samples and more centers. It is also necessary to ensure high quality and safe delivery for all childbearing women. In addition, maternal age, height, BMI, weight gain during pregnancy, gestational age, mode of labor induction, meconium-stained amniotic fluid, presence of complications, neonatal weight, and neonatal gender affected the delivery mode.

Author contributions

All authors contributed to the creation of the article as follows: BH had the original idea of the study. FCS, MHS, YGC and FFW contributed to data collection. FCS and MHS analyzed the data. FCS wrote the first draft of the paper. BH and FCS provided critical review and interpretation of the results. BH provided professional language. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The protocol for this study was approved by The Institutional Review Board at this center (2018019). As this study is a retrospective study, all individuals participating in the study are exempted from informed consent.

Conflict of interest

The authors declare no conflict of interest.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at https://www.imrpress.com/journal/CEOG/49/3/10.31083/j.ceog4903076.

References

[1] Sandall J, Tribe RM, Avery L, Mola G, Visser GH, Homer CS, *et al*. Short-term and long-term effects of caesarean section on the health of women and children. The Lancet. 2018; 392: 1349–1357.

[2] Tribe RM, Taylor PD, Kelly NM, Rees D, Sandall J, Kennedy HP. Parturition and the perinatal period: can mode of delivery impact on the future health of the neonate? The Journal of Physiology. 2018; 596: 5709–5722.

[3] Hosseini Tabaghdehi M, Kolahdozan S, Keramat A, Shahhossein Z, Moosazadeh M, Motaghi Z. Prevalence and factors affecting the negative childbirth experiences: a systematic review. The Journal of Maternal-Fetal & Neonatal Medicine. 2020; 33: 3849–3856.

[4] Tolcher MC, Holbert MR, Weaver AL, McGree ME, Olson JE, El-Nashar SA, *et al*. Predicting Cesarean Delivery after Induction of Labor among Nulliparous Women at Term. Obstetrics & Gynecology. 2015; 126: 1059–1068.

[5] Levine LD, Downes KL, Parry S, Elovitz MA, Srinivas SK. 468: Predicting cesarean delivery for women undergoing an induction of labor with an unfavorable cervix. American Journal of Obstetrics and Gynecology. 2017; 216: S277.

[6] Burke N, Burke G, Breathnach F, McAuliffe F, Morrison JJ, Turner M, *et al*. Prediction of cesarean delivery in the term nulliparous woman: results from the prospective, multicenter Genesis study. American Journal of Obstetrics and Gynecology. 2017; 216: 598.e1–598.e11.

[7] Levine LD, Downes KL, Parry S, Elovitz MA, Sammel MD, Srinivas SK. A validated calculator to estimate risk of cesarean after an induction of labor with an unfavorable cervix. American Journal of Obstetrics and Gynecology. 2018; 218: 254.e1–254.e7.

[8] Rossi RM, Requarth EW, Warshak CR, Dufendach K, Hall ES, DeFranco EA. Predictive Model for Failed Induction of Labor among Obese Women. Obstetrics & Gynecology. 2019; 134: 485–493.

[9] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143: 29–36.

[10] Li H, Luo S, Trasande L, Hellerstein S, Kang C, Li J, *et al*. Geographic Variations and Temporal Trends in Cesarean Delivery Rates in China, 2008–2014. The Journal of the American Medical Association. 2017; 317: 69–76.

[11] Kawakita T, Reddy UM, Huang CC, Auguste TC, Bauer D, Overcash RT. Predicting Vaginal Delivery in Nulliparous Women Undergoing Induction of Labor at Term. American Journal of Perinatology. 2018; 35: 660–668.

[12] Verhoeven CJM, Oudenaarden A, Hermus MAA, Porath MM, Oei SG, Mol BWJ. Validation of models that predict Cesarean section after induction of labor. Ultrasound in Obstetrics and Gynecology. 2009; 34: 316–321.

[13] Kwon HY, Kwon J, Park YW, Kim Y. The risk of emergency cesarean section after failure of vaginal delivery according to prepregnancy body mass index or gestational weight gain by the 2009 Institute of Medicine guidelines. Obstetrics & Gynecology Science. 2016; 59: 169–177.

[14] Grobman WA, Rice MM, Reddy UM, Tita ATN, Silver RM, Mallett G, *et al*. Labor Induction versus Expectant Management in Low-Risk Nulliparous Women. New England Journal of Medicine. 2018; 379: 513–523.

[15] Alavifard S, Meier K, D'Souza R. Prediction calculator for induction of labor: no Holy Grail yet! American Journal of Obstetrics and Gynecology. 2018; 219: 419–420.

[16] Lipschuetz M, Cohen SM, Israel A, Baron J, Porat S, Valsky DV, *et al*. Sonographic large fetal head circumference and risk of cesarean delivery. American Journal of Obstetrics and Gynecology. 2018; 218: 339.e1–339.e7.

[17] Kleinrouweler CE, Cheong-See FM, Collins GS, Kwee A, Thangaratinam S, Khan KS, *et al*. Prognostic models in obstetrics: available, but far from applicable. American Journal of Obstetrics and Gynecology. 2016; 214: 79–90.e36.