

EVALUATING THE ACCURACY OF THE MODIFIED BISHOP SCORE IN PREDICTING DELIVERY OUTCOMES AMONG TERM PREGNANCIES UNDERGOING LABOR INDUCTION

Dr. QURAT UL AIN KHALIL*

Post Graduate Trainee, FCPS II Gynecology & Obstetrics, Rawalpindi Teaching Hospital, Rawalpindi.
*Corresponding Author Email: ainikhalil91@gmail.com

Dr. SAIMA FAREED

Post graduate trainee, FCPS II Gynecology & Obstetrics, Rawalpindi Teaching Hospital, Rawalpindi.
Email: drrsaimakhan@gmail.com

Dr. SOBIA NAWAZ MALIK

Associate Professor, Department of Gynecology & Obstetrics, Rawalpindi Medical University, Rawalpindi.
Email: snmgynobs@gmail.com

Abstract

Aim: To evaluate the diagnostic performance of the Modified Bishop Score (MBS) in predicting successful vaginal delivery among term primigravida women undergoing labor induction. **Methods:** This prospective observational study was conducted at a tertiary care facility and included 130 term primigravida women induced with dinoprostone. Prior to induction, the MBS was recorded for each participant. Patients were followed until the mode of delivery was determined. Statistical analysis was performed using SPSS version 26, with diagnostic metrics including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy. **Results:** An MBS threshold of ≥ 6 demonstrated strong predictive value for vaginal delivery, with sensitivity of 73.2%, specificity of 89.6%, PPV of 92.3%, NPV of 66.2%, and overall accuracy of 79.2%. ROC curve analysis confirmed the score's discriminative ability. Additionally, higher maternal age and body mass index were significantly associated with increased rates of cesarean section. **Conclusion:** The Modified Bishop Score is a practical, reliable, and cost-effective tool for predicting delivery outcomes in induced term pregnancies. Its high diagnostic accuracy supports its use in clinical decision-making, particularly in resource-constrained settings.

Keywords: Modified Bishop Score, Induction of Labor, Vaginal Delivery, Cesarean Section, Cervical Favorability.

INTRODUCTION

The term induction of labor (IOL) is defined as the artificial induction of uterine contractions through mechanical or pharmacological methods prior to the spontaneous onset of labor and the goal is to deliver the fetus via the vagina post the stage of fetal viability and intact membranes and absence of spontaneous labor signals [1]. Obtaining the indication of labor induction has been rising around the world, demonstrating the changing trends of obstetrics and the change of the idea of preventing complications of post-term and high-risk pregnancies. It has been reported that about 9.5-33.7% of total pregnancies in cases of labor induction have been reported and in the developed nations have been reported as high as 25 percent whereas in Sri Lanka and 4.5 percent in Niger are the lowest [2-5]. Even though there are limited national data, some tertiary care

facilities indicate an induction rate of 40 percent [3]. Primigravidae are also more likely to induce as compared to multigravidae, and their rates of cesarean delivery after induction are higher (49.8%), because of slower cervical ripening and poor uterine responsiveness [2]. IOL is determined by a number of factors that are maternal and fetal, the most important one being the cervical readiness, normally determined by the Bishop Score. This system was to be used to assess cervical position prior to artificial rupture of membranes (ARM) (6).

The bishop score has been the most frequently used tool in determining cervical readiness to labor by measuring such factors as cervical dilatation, effacement, position, consistency, and the fetal station. Nevertheless, in spite of its popularity, it has frequently been criticized to be inconsistent among observers and its inefficacy in being able to forecast labor outcomes accurately. It has been reported that its ability to predict a successful vaginal delivery may be as low as 23% up to 64% [1]. In order to overcome these shortcomings a Modified Bishop Score (MBS) was proposed replacing cervical effacement with a quantifiable cervical length and increasing the maximum total score to 13 points.

The change does not change the criteria of evaluating fetal station, but instead tries to make the assessment more objective and clinically helpful [7]. Scientists have since endeavored in trying to establish the most reliable cutoff point in forecasting positive induction results. Although the old Bishop score indicates a score of over 8 to indicate a positive cervix, a number of studies have suggested that a score of 5 or 6 on the MBS can be a better indicator of successful vaginal delivery particularly in term pregnancies and in multiple pregnancies [8,9]. As an example, Sinha et al. observed that the sensitivity of the MBS with a cutoff of 6 was 28.9% and specificity of 95%, which suggests a successful induction, which is why the modified version can be of clinical value [9]. However, predictive accuracy of the MBS still differs in various populations and hospital environments. The inconsistency is further augmented by the differences in the clinical performance and interpretation of digital cervical examinations by the clinicians, as the scoring of the test can be affected by personal experience and technique.

This inter-observer variation remains a significant limitation in labor assessment. Therefore, ongoing evaluation of the diagnostic accuracy of the MBS is important, particularly in resource-limited healthcare environments where physicians often rely on clinical judgment to guide management and make timely decisions for both maternal and fetal safety. The present study aims to assess the diagnostic accuracy of the MBS in predicting successful vaginal delivery (gold standard) among women undergoing induction of labor at term. Specifically, it seeks to determine whether MBS can reliably forecast the likelihood of successful vaginal delivery, thereby helping clinicians optimize patient selection, reduce unnecessary cesarean sections, and enhance maternal and fetal outcomes. Validating this scoring system in a local population may also contribute to standardizing induction practices and improving decision-making in obstetric care.

MATERIALS AND METHODS

The study was conducted in the Department of Obstetrics and Gynecology, Rawalpindi Teaching Hospital, Rawalpindi, over a duration of six months. A total of 130 patients were enrolled using a non-probability consecutive sampling technique, based on a sample size calculation with a 95% confidence level, expected sensitivity of 0.29, specificity of 0.95, prevalence of 0.62, and desired precision of 0.10. This cross-sectional validation study included primigravida women with singleton, live fetuses in cephalic presentation at term (37–40+6 weeks of gestation) in whom induction of labor was clinically indicated, such as in cases of gestational hypertension, gestational diabetes, or oligohydramnios. Patients with contraindications to vaginal delivery, including previous uterine surgery, placenta previa, placental abruption, ruptured membranes, or severe medical disorders such as eclampsia, preeclampsia, or cardiovascular disease, as well as those with uncertain gestational age without an early ultrasound, were excluded. Informed consent was secured from all participants prior to data collection. Each patient underwent detailed history taking, clinical examination, and cervical assessment using the Modified Bishop Scoring (MBS) system, which provided scores ranging from 0 to 13 based on cervical dilation (0–>4 cm), effacement (0–≥75%), fetal station (-3 to +3), cervical consistency (firm, medium, soft), and position (posterior, mid, anterior). A score of ≤6 was considered unfavorable for spontaneous labor and indicated the need for induction. Induction was performed using Tablet Prostin E2 (dinoprostone 3 mg) inserted vaginally, with two doses administered six hours apart. After each dose, patients were monitored for uterine contractions and changes in the bishop score. If cervical improvement was inadequate after six hours, re-induction was performed, and further assessments were carried out at six-hour intervals. If no improvement was achieved within 24 hours, induction was deemed failed, and cesarean section was considered. Data recorded included maternal age, BMI, gestational age, time to delivery, type of delivery, requirement for further intervention, and maternal or fetal complications. The accuracy of the MBS was defined as its ability to correctly predict delivery outcomes (vaginal delivery or cesarean section) based on the cervical score. Data was analyzed using SPSS v.26, where continuous (such as age, BMI, and gestational age) variables were expressed as mean ± SD, while categorical (like indication for induction and delivery outcomes) variables were presented as frequencies and percentages. Effect modifiers, including age, BMI, and gestational age, were controlled through stratification, and post-stratification analysis was performed using 2×2 contingency tables to compute sensitivity, specificity, PPV, NPV, and overall diagnostic accuracy. Receiver Operating Characteristic (ROC) curves were plotted to assess the predictive performance of the MBS, and a p-value ≤0.05 was considered statistically significant.

RESULTS

A total of 130 pregnant women undergoing induction of labor at term were included in the study. The mean age of the participants was 27.28 ± 3.99 years, while the mean gestational age was 38.95 ± 1.04 weeks. The mean Body Mass Index (BMI) was 25.23 ±

3.77 kg/m², indicating that most participants fell within the normal to overweight categories. Regarding BMI classification, 50.0% of women had normal weight. The most common indication for induction was intrauterine growth restriction (IUGR) or fetal concern, followed by post-dated pregnancies, gestational hypertension, gestational diabetes, and oligohydramnios. Out of 130 patients, 63.1% achieved vaginal delivery, whereas 36.9% underwent cesarean section. The mean MBS at admission was 6.05 ± 3.36 , and half of the participants had favorable scores, while the remaining half had unfavorable scores (Table 1).

Table 1: Descriptive Characteristics of the Study Population

Variable	Category	Frequency (%)	Mean $\pm$ SD
Age (years)			27.28 $\pm$ 3.99
Gestational Age (weeks)		-	38.95 $\pm$ 1.04
Body Mass Index (BMI)		-	25.23 $\pm$ 3.77
	Underweight	1 (0.8)	-
	Normal weight	65 (50.0)	-
	Overweight	47 (36.2)	-
	Obese	17 (13.1)	-
Indications for Induction	IUGR / Fetal concern	35 (26.9)	-
	Post-dates (≥ 41 weeks)	33 (25.4)	-
	Gestational hypertension	23 (17.7)	-
	Gestational diabetes	20 (15.4)	-
	Oligohydramnios	19 (14.6)	-
Final Delivery Outcome	Vaginal delivery	82 (63.1)	-
	Cesarean section	48 (36.9)	-
Modified Bishop Score (MBS)		-	6.05 $\pm$ 3.36
	< 6 (Unfavorable)	65 (50.0)	-
	≥ 6 (Favorable)	65 (50.0)	-

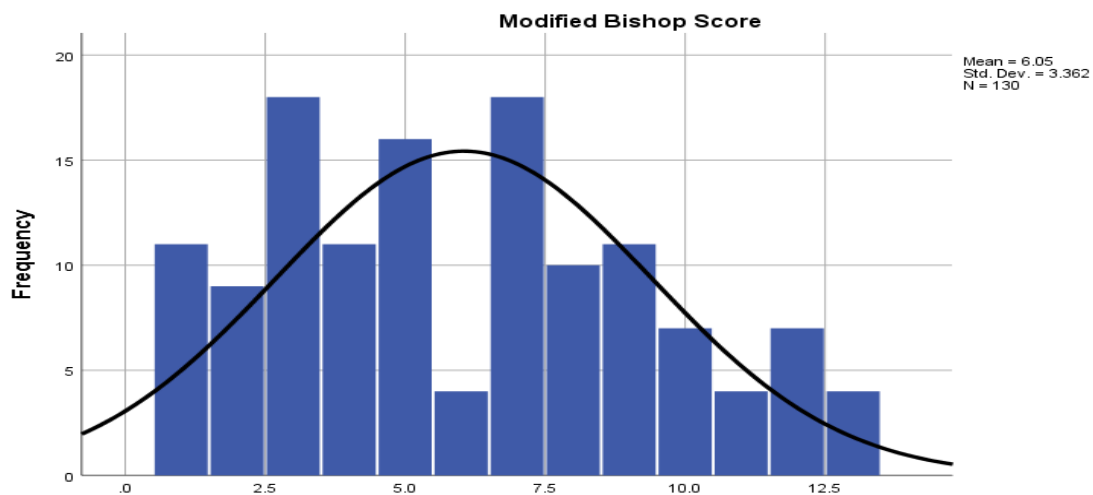


Figure 1: Distribution of Modified Bishop Score among Study Participants

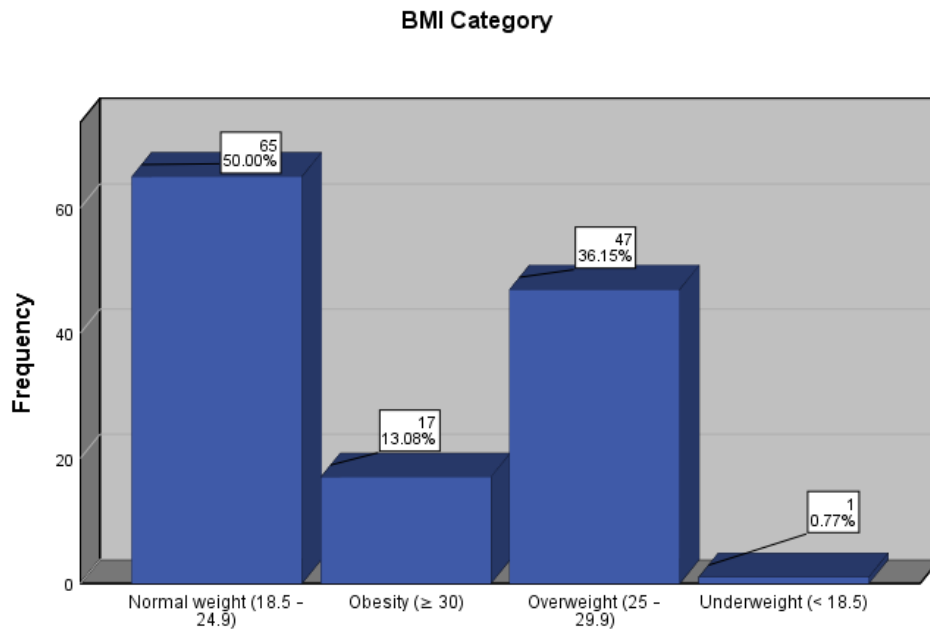


Figure 2: Distribution of Participants According to BMI Category

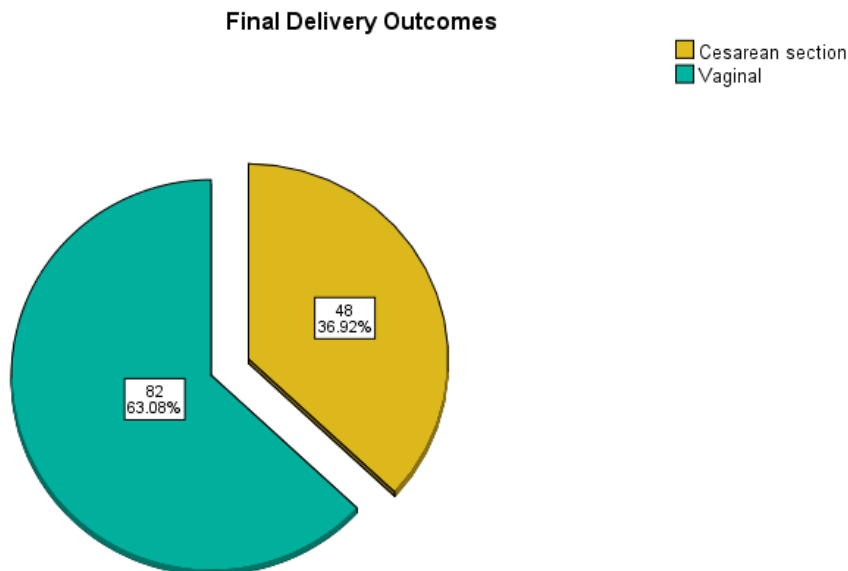


Figure 3: Final Delivery Outcomes among Study Participants

To evaluate the association of maternal factors (age, gestational age, and BMI) with delivery outcomes, an ANOVA was performed. The results showed statistically significant differences across groups for all three continuous variables. Age had a significant association with delivery outcomes ($F = 10.937$, $p = 0.001$), indicating that older maternal

age was linked to a higher likelihood of cesarean section. Gestational age was also significantly different between groups ($F = 29.456$, $p < 0.001$), suggesting that women with lower gestational ages were more likely to undergo cesarean section. Similarly, BMI demonstrated a significant effect ($F = 20.444$, $p < 0.001$), showing that higher BMI was associated with increased cesarean rates. Post-stratification was performed to control for potential confounders such as age, BMI, and gestational age. The results showed significant associations between these variables and the mode of delivery. A strong relationship ($p < 0.001$) was found, with higher cesarean rates among obese women (27/48; 56.2%) compared to normal-weight participants. Women aged >30 years were more likely to require cesarean section (20/33; 60.6%) compared to those aged ≤ 25 years, where vaginal delivery predominated (31/34; 91.2%). Those at ≤ 38 weeks were more prone to cesarean section (12/22; 54.5%), whereas higher gestational ages were associated with vaginal deliveries ($p = 0.016$) (Table 2).

Table 2: Post-Stratification by Maternal Characteristics and Delivery Outcome

Variable	Category	Vaginal Delivery n (%)	Cesarean n (%)	p-Value
BMI Category	Underweight	1	0	.000
	Normal	61	4	
	Overweight	0	17	
	Obese	20	27	
Age Group (years)	≤ 25	31	3	.000
	25–30	38	25	
	>30	13	20	
Gestational Age (weeks)	≤ 38	10	12	.016
	39–40	38	27	
	>40	34	9	

The diagnostic accuracy of the MBS in predicting successful vaginal delivery was evaluated using 2×2 contingency analysis. Among the 130 participants, 65 had favorable MBS (≥ 6), and 65 had unfavorable scores (< 6). Out of those with favorable scores, 60 (True Positives) achieved vaginal delivery, and 5 (False Positives) required cesarean section. Conversely, 22 (False Negatives) with unfavorable scores had vaginal delivery, while 43 (True Negatives) underwent cesarean section.

Table 3: Diagnostic Accuracy of Modified Bishop Score (MBS)

MBS (Test Result)	Vaginal Delivery (Positive)	Cesarean Section (Negative)	Total	P value
Positive (≥ 6)	60 (TP)	5 (FP)	65	.000
Negative (< 6)	22 (FN)	43 (TN)	65	
Total	82	48	130	
Metric		Formula	Result	
Sensitivity		$TP / (TP + FN)$	$60 / (60 + 22) = 73.17\%$	
Specificity		$TN / (TN + FP)$	$43 / (43 + 5) = 89.58\%$	
PPV		$TP / (TP + FP)$	$60 / (60 + 5) = 92.31\%$	
NPV		$TN / (TN + FN)$	$43 / (43 + 22) = 66.15\%$	
Diagnostic Accuracy		$(TP + TN) / N$	$(60 + 43) / 130 = 79.23\%$	

ROC Curve for Vaginal Delivery (Positive Outcome)

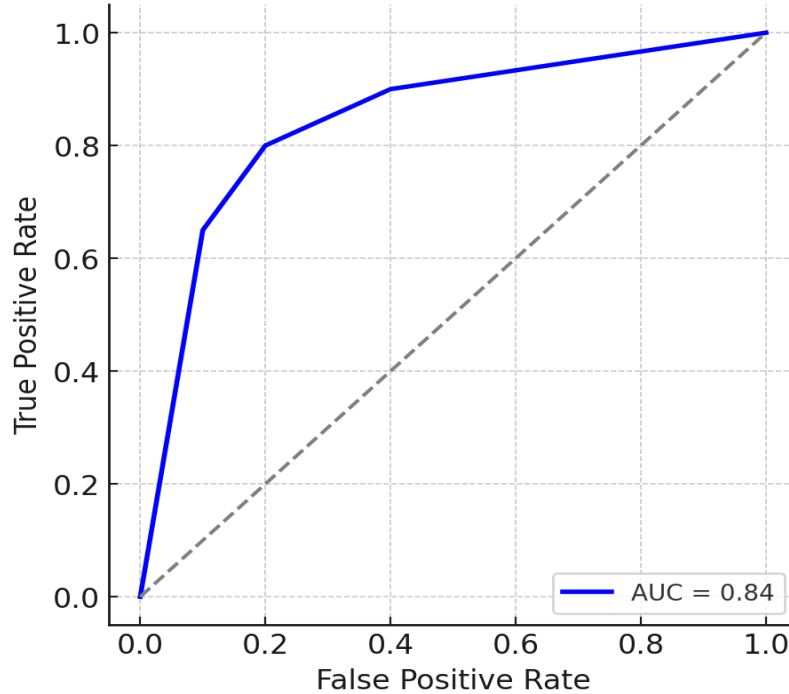


Figure 4: The ROC curve demonstrates good discriminatory ability of the model to distinguish vaginal delivery as the positive outcome.

DISCUSSION

In this study involving 130 first-time mothers at term who underwent labor induction with dinoprostone, the MBS with a cutoff value of > 6 showed good reliability in predicting delivery outcomes. The score demonstrated a sensitivity of 73.2%, specificity of 89.6%, and an overall accuracy of 79.2%. These results indicate that the MBS is particularly effective in identifying women who are likely to achieve a successful vaginal delivery. The ROC curve also confirmed that the score has strong ability to distinguish between those who will deliver vaginally and those who may require a cesarean section. The high specificity and PPV found in this study support previous research showing that when the cervix is favorable according to the Bishop Score, the likelihood of a successful induction and vaginal birth is high. Overall, the MBS remains a simple, practical, and valuable tool for guiding clinical decisions during labor induction.

The modest sensitivity and NPV reflect the long-known limitation that some women with apparently unfavorable cervixes still respond well to induction. In a retrospective study of 91 primigravidae, those with MBS > 6 had a vaginal delivery rate of 78.2%, yielding sensitivity $\sim 78.2\%$ and specificity $\sim 50\%$, with PPV $\sim 97.1\%$ and NPV $\sim 9.5\%$ (i.e., among those scoring < 6 , very few failed vaginal delivery) [14].

The lower specificity in that study may reflect different induction protocols, heterogeneous population, or sample size effects. Still, the high PPV in that study aligns with your observation: favorable MBS strongly predicts vaginal delivery. In the recent study by Milatović et al. comparing pre-induction ultrasound parameters and Bishop score, both cervical length and the digital Bishop score were independent predictors of successful induction (vaginal delivery), and adding ultrasound parameters (like funneling and posterior cervical angle) improved prediction ($p < 0.001$) [1,15].

Their findings support the concept that combining ultrasound measures and clinical scoring might outperform either alone. In this study, an MBS of ≥ 6 already showed strong performance; supplementing it with ultrasound might further boost sensitivity or assist in borderline cases. Tripathy et al. (2024) argued that while Bishop score and transvaginal ultrasound both predict induction success, TVS measurements are more objective and less prone to interobserver variability, particularly in ambiguous cases [4]. Another comparative study in Pakistan by Iftikhar et al. (2023) involving 520 primigravidae concluded that both the bishop score and TVS were effective, but combining both gave additive predictive power, improving both predictive values by over 10% each [16]. Antonios et al. (2025) reviewed evidence and affirmed that while the bishop score is widely used owing to its simplicity, it suffers from subjectivity, interexaminer variability, and declining predictive power in certain populations (such as nulliparas or inductions for non-postdate reasons).

They recommended combining Bishop scoring with other modalities (e.g., ultrasound, clinical predictors) or newer models to optimize outcomes [17]. In older, larger studies, the original Bishop score has also been validated. Vroenenraets et al. (2005) studied over 1,300 women and found that a Bishop score ≤ 5 was a significant risk factor for cesarean delivery (adjusted OR ~ 2.32) in induction settings, even after controlling for confounders such as maternal age, BMI, and fetal size [18]. Other studies also highlight the induction-cesarean dynamics. For example, Bashir et al. in Pakistan analyzed primigravidas at term and reported frequencies and risk factors for cesarean after induction, though without formal diagnostic accuracy analysis [2].

Similarly, Khattak et al. (2023) reported the prevalence of cesarean following induction with unfavorable Bishop scoring, reinforcing that in Pakistani populations, induction failure remains a clinical challenge [3]. This study observed that **higher BMI and older maternal age** were significantly associated with higher cesarean rates after induction. These results align with international evidence: for instance, Milatović et al. (2024) found that obesity increased the risk of cesarean after induction, with incremental risk across BMI categories [1]. Likewise, many obstetric registries (Nordic registers) show that advanced maternal age is independently associated with increased rates of cesarean even when adjusting for induction status [17]. These effect modifiers underscore that while MBS is predictive, its performance is not uniform across all patient subgroups. In higher BMI or older age strata, the threshold for expecting success might be shifted upward, or clinicians might be more cautious.

Thus, “favorable cervix” should be interpreted in the context of patient risk profile rather than absolute score alone. A strength of recent work is focusing on primigravidae and using a clear induction protocol. Moreover, study employed post-stratification to control for confounders and used ROC analysis to quantify discrimination. However, several limitations merit acknowledgment.

First, the sample is from a single center, limiting external generalizability, especially to multiparous women or different ethnic populations. Second, digital examination of cervical parameters is subject to interobserver variability; even with training, variation in scoring may affect sensitivity/specificity. Third, the absence of adjunctive ultrasound or biochemical markers in this study constrains direct comparisons to combined models in the literature. Clinically, these findings support the continued use of the MBS (≥ 6) as a practical tool to guide induction decisions in term, primigravida patients.

The high PPV and specificity suggest that when the cervix is favorable, one can proceed with reasonable confidence in induction success. However, due to moderate sensitivity and NPV, a low score should not automatically exclude induction; rather, it should prompt more vigilant monitoring, adjunctive assessments (e.g., ultrasound), or individualized decision-making.

Future research should aim to validate these findings in multicenter cohorts, include multiparous women, compare MBS with ultrasound models prospectively, and explore whether dynamic changes in MBS after ripening doses add incremental predictive value. Integration into clinical decision algorithms or mobile app tools might also improve uptake and utility.

CONCLUSION

Study concluded that a Modified Bishop Score ≥ 6 has strong specificity and positive predictive value for predicting vaginal delivery after induction of labor at term in primigravidae. Its sensitivity and negative predictive value are more modest, reflecting that not all “unfavorable” cervixes fail induction. Stratified analyses highlight that maternal BMI and age modulate induction success. While MBS remains a valuable tool, its predictive performance can likely be enhanced by combining with ultrasound or newer techniques. These insights inform clinical selection and counseling, particularly in resource-constrained settings, and offer a platform for further validation and refinement.

Author Contribution

Dr. Qurat UL Ain Khalil - Principal Author Concept, Drafting, Data Acquisition, Data analysis & Interpretation

Dr. Saima Fareed - Author Concept & design, Critical review for important intellectual content, Data analysis & Interpretation

Dr. Sobia Nawaz Malik - Author Final approval of the version to be published

Acknowledgements

None.

References

- 1) Milatović S, Krsman A, Baturan B, Dragutinović Đ, Ilić Đ, Stajić D. Comparing Pre-Induction Ultrasound parameters and the bishop score to determine whether labor induction is successful. *Medicina*. 2024 Jul 12;60(7):1127. Available from: <https://doi.org/10.3390/medicina60071127>
- 2) Bashir et al. Frequency and Factors Leading to Caesarean Section after Induction of Labour in Primigravidas at Term. *Pakistan Journal of Medical & Health Sciences*. 2021;15(9).
- 3) Khattak KH, Furqan S, Bugti S, Javed S, Bibi S, Shamsher S. Prevalence of Delivery Through C-Section Followed by the Induction of Labour in Nulliparous Patients Showing Undesirable Bishops Score. *Pakistan Journal of Medical & Health Sciences*. 2023 May 12;17(03):521-.
- 4) Tripathy P, Pati T, Baby P, Mohapatra SK. Prevalence and predictors of failed induction. *Int J Pharm Sci Rev Res*. 2016;39(2):189-94.
- 5) Lueth GD, Kebede A, Medhanyie AA. Prevalence, outcomes and associated factors of labor induction among women delivered at public hospitals of MEKELLE town- (a hospital-based cross-sectional study). *BMC pregnancy and childbirth*. 2020 Dec; 20:1-0.
- 6) Pirzada H, Shabbier N, Ara I, Hussain S, Akram R, Khokhar S. Comparison of Transvaginal Ultrasound Cervical Length with Bishop Score in Predicting Cesarean Section after Labor Induction. *Pakistan Journal of Health Sciences*. 2024 Aug 10;125–31. Available from: <https://doi.org/10.54393/pjhs.v5i07.1695>
- 7) Wormer KC, Bauer A, Williford AE. Bishop Score. 2024 Jul 17. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2025 Jan–. PMID: 29261961.
- 8) Inde Y, Nakai A, Sekiguchi A, Hayashi M, Takeshita T. Cervical dilatation curves of spontaneous deliveries in pregnant Japanese females. *International Journal of Medical Sciences*. 2018 Jan 1;15(6):549–56. Available from: <https://doi.org/10.7150/ijms.23505>
- 9) Sinha A, Sinha M, L, Sharma S, Prasad D, Goel Kumari M. et al. A comparative study between modified bishop score and transvaginal sonography to predict successful induction. *Family Med Prim Care* 2024;13;4438-43
- 10) Ali BN, Almonaem HS, Yousef FS. Comparison Between Manipal Scoring System and Modified Bishop scoring in predicting successful labour induction. *Al-Azhar International Medical Journal*. 2024;5(3):31.
- 11) Li SF, Ju HH, Feng CS. Effect of cervical Bishop score on induction of labor at term in primiparas using Foley catheter balloon: a retrospective study. *BMC Pregnancy and Childbirth*. 2024 May 31;24(1):401.
- 12) National Child and Maternal Health Education Program. [Last reviewed June 2 2022]. Full term has a new meaning. URL: <https://www.nichd.nih.gov/ncmh/ep/initiatives/know-your-terms/moms>
- 13) Negrini R, da Silva Ferreira RD, Guimarães DZ. Value-based care in obstetrics: comparison between vaginal birth and cesarean section. *BMC Pregnancy and Childbirth*. 2021 Apr 26;21(1):333.
- 14) Galzie SK, Rao SB. Cervical effacement, as an independent parameter versus modified bishop score, for predicting the favorability of vaginal delivery in a primigravida at 40 weeks gestation and beyond. *Int J Reprod Contracept Obstet Gynecol* [Internet]. 2020 Apr. 28 [cited 2025 Oct. 9];9(5):1831-7. Available from: <https://www.ijrcog.org/index.php/ijrcog/article/view/8315>
- 15) Rācene L, Kīse L, Pitkēviča I, Rostoka Z, Sārta B, Priedniece M, Vecvagare A, Lapidus L, Kīvīte-Urtāne A, Rezeberga D, Vedmedovska N. The significance of ultrasound parameters and clinical factors in predicting successful labor induction among nulliparous women. *The Journal of MaTernal-*

feTal & neonaTal Medicine. 2025 Dec 31;38(1):2450405.

- 16) Iftikhar T, Qazi N, Malik N, Hamid N, Gul N, Rauf R. Prediction of successful induction of labour jointly using bishop score and transvaginal sonography in primigravida women in pakistan. Annals of PIMS-Shaheed Zulfiqar Ali Bhutto Medical University. 2023 May 31;19(2):141-6.
- 17) Antonios M, Zacharias F, Ekaterini D, Daskalakis G, Panagiotis A. Role of the Bishop Score in Predicting Successful Induction of Vaginal Delivery: A Systematic Review of Current Evidence. Cureus. 2025;17(7).
- 18) Vrouwenraets FP, Roumen FJ, Dehing CJ, Van den Akker ES, Aarts MJ, Scheve EJ. Bishop score and risk of cesarean delivery after induction of labor in nulliparous women. Obstetrics & Gynecology. 2005 Apr 1;105(4):690-7. Doi: 10.1097/01.AOG.0000152338.76759.38