Comparison of transvaginal ultrasonographic and digital cervical assessment in predicting successful induction of labour in nulliparous pregnancy beyond 40 weeks with unfavourable cervix – A prospective cohort study

Ransiri PADL a, Rathnayake RMCJ b, Kandauda IC c

Abstract

Introduction: To date, the Bishop score remains as the standard method to predict the success of labour induction. Possible role of transvaginal ultrasound measurement of cervical length provides conflicting evidences. This study compares the value of transvaginal ultrasound cervical length (TVSCL) and the Bishop score (BS), in predicting the success of labour induction in nulliparous pregnancy beyond 40 weeks with unfavourable cervix.

Method: In this prospective cohort study pre induction TVSCL and Bishop score were measured in 392 nulliparous women who underwent induction with vaginal Prostaglandin (PGE2) at 40 weeks+6 days. Achieving a cervical dilatation of 8cm was considered as a successful induction. Predictive value of TVSCL and BS, in determining successful induction, and amniotomy to successful induction time interval (TI) were analyzed using chi-square test, unpaired t-test, multiple logistic regression, Pearson’s co-efficient and receiver-operating characteristics (ROC) curves.

Results: Induction of labour was successful in 75.5% (n=296) of the women. There was a significant difference in mean TVSCL, between successful and failed induction groups (P=0.02). Best cut-off value of TVSCL for predicting successful induction was 3.3cm. However TVSCL failed to demonstrate significant discriminatory value (Area under the ROC curve (AUC)=0.545;95% CI,0.496-0.597;P=0.17). Meanwhile the AUC for the Bishop score >3 was significant (AUC=0.548;95%CI,0.548-0.647;P=0.006). However, sensitivity and specificity of the Bishop score in predicting induction success were 76.0% and 44.8% respectively. There was no significant association between TVSCL and TI (R=0.02, P=0.06).

Conclusion: TVSCL is not an accurate predictor for the outcome of labour induction in nulliparous pregnancy beyond 40 weeks with unfavourable cervix. Nevertheless, the Bishop score appears to be of poor predictive value.

Key words: Induction of labour, Transvaginal ultrasonographic cervical length

INTRODUCTION

Induction of labour is a relatively common intervention in modern obstetric practice. WHO Global Survey on maternal and perinatal health involved in 24 countries revealed an average labour induction rate of 12.1% in Asia and, Sri Lanka had the highest rate of labour induction (35.5%).

The clinical requirement for induction of labour arises from circumstances in which it is believed that the outcome of the pregnancy will be better if it is artificially interrupted rather than being left to follow its natural course. Prolonged pregnancy is one such circumstance when induction is strongly recommended, as it is associated with a higher perinatal mortality, morbidity rate and a higher risk of complications during delivery. At 41 weeks gestation, approximately 9% of women remain undelivered.

As similar to any other clinical intervention induction of labour associated with several inherent fetal and maternal complications. Overall, it is estimated that failed induction in the presence of an unfavourable cervix is found in 15% of cases. Uterine hyperstimulation is a fairly common fetal complication and has an incidence of 1-5%. Therefore a close fetal monitoring is recommended during the induction process. In addition, it will take 12-24 hours to complete the procedure and also produces a greater analgesic requirement than does spontaneous labour. The cost for induction agents also represents a substantial percentage of annual drug estimates. As a result, induction of labour place more strains on labour wards than spontaneous labour. Therefore, it is important to predict...
successful induction of labour reliably and thus determine whether it would be better to carryout caesarean section rather than attempt to induce labour.

The Bishop score was originally introduced in 1964 as a method to predict the time interval from vaginal examination until the onset of spontaneous labour in multiparous women. Subsequently, the Bishop score was used as a method of assessing the favourability of the cervix in women undergoing induction of labour. The score was later modified by Burnett. To date, the Bishop score remain as the standard method to predict the duration and, successfulness of induced labour. However, digital cervical assessment using the Bishop score has certain limitations. It is subjective, imprecise and has wide intra- and inter-observer variations. Furthermore Bishop score may be of limited value in women with a lower score.

Two other methods for predicting successful labour induction have been investigated recently: screening for fetal fibronectin in vaginal secretions and transvaginal sonography of uterine cervix. Studies on fetal fibronectin consistently indicated that the fetal fibronectin in vaginal secretions is a reliable marker in predicting successful labour induction despite of its high cost.

In recent years, numerous studies have evaluated the possible role of transvaginal ultrasound examination of cervical length in the prediction of labour induction outcomes compared to Bishop score and provide conflicting results. Some studies have found a significant association between cervical length measured by transvaginal ultra-sound scan (TVSCL) and successful labour induction and also demonstrated a better performance than Bishop score in predicting the successful induction specially in situations with low Bishop score. Meanwhile another set of studies have found totally opposite results. However transvaginal ultrasonography is more objective and more reproducible than Bishop score. It has also been shown to have reduced intra- and inter-observer variability and transvaginal sonogram is also more accurate due to its ability in measuring supravaginal part of the cervix.

Therefore our primary objective is to assess value of ultrasonographically measured cervical length, in comparison to the Bishop score, in predicting the successful labour induction and with a secondary objective of predicting time interval from amniotomy to successful induction (TI) in a population of nulliparous pregnancy beyond 40 weeks with low Bishop score.

**MATERIALS AND METHODS**

This prospective cohort study was conducted at the University Obstetrics Unit, Teaching Hospital, Peradeniya, Sri Lanka from July 2011 to April 2012 among primigravid women who admitted with pregnancy beyond 40 weeks and who fulfilled the following recruitment criteria. The inclusion criteria were: nulliparous women, gestational age 40 weeks+ 6 days (confirmed by ultrasound scan measurements before 16 weeks of gestation according to NICE guideline), singleton pregnancy, cephalic presentation, intact membrane, Bishop score less than or equal to 5, and who given the informed consent.

The exclusion criteria were; maternal age less than 18 years, any maternal or pregnancy related comorbid conditions like gestational diabetes, pregnancy-induced hypertension, placenta previa, any fetal abnormalities like congenital abnormality, intrauterine growth retardation and mothers who had any surgical procedure involving uterine body, cervix or pelvis. To obtain maximum number of samples to achieve objectives, a preliminary assumption was made that the proportion of successful induction predicted by either method was 50%. A formula as described by Lwanga and Lemeshow was used to calculate the sample size with confidence level at 95% and precision at 5%. A 10% drop out rate was also added and a total of 428 women were recruited.

Being a South East Asian population, according to our unit policy all the women with uncomplicated pregnancies offered cervical ripening at the gestational age of 40 weeks+ 6 days if the cervix is unfavourable. After admitting to the unit, digital vaginal examination was performed and Bishop score was calculated as recommended by NICE clinical guideline 70, the lowest possible Bishop score being 0 and the maximum possible score being 12. If the Bishop score was equal to or less than 5, the status of the cervix was considered as unfavourable and after receiving informed consent the woman was recruited to the study. Immediately after digital vaginal examination the woman was directed to the transvaginal ultrasonographic cervical assessment. (TVS)

TVS of the cervix was done using a ultrasound system equipped with a 6.5MHz micro-convex device, by an experienced operator who was blinded to the digital examination findings. A standardized technique was used to measure the cervical length. After the patient emptied her bladder, imaging was carried out in the dorsal lithotomy position. Recommended steps taken to avoid undue pressure on the cervix. The whole length of the sonoluent endocervical mucosa was identified in the sagital section, and the image was magnified so that this occupied 75% of the screen. Callipers were placed from the external os to the V-shaped indentation marking of the internal os, and this distance was measured in a straight line. Three measurements were made over a period of 3 minutes to allow any change in the state of the cervix, and the shortest measurement was reported. During the rest of the induction process all the persons involving decision making were blinded to the TVS findings and all the decisions were taken according the digital cervical assessment findings.

If the Bishop score was 5 or less the induction of labour was started within one hour after the cervical assessment. Dinoprostone (PGE) mg vaginal tablet was inserted and time of the insertion was recorded. Fetal and maternal monitoring was done according to the NICE clinical guideline 70. If labour was not commenced spontaneously the
women was reassessed 8 hours later. If she did not exhibit regular uterine contraction and Bishop Score was less than 6, a second dose of PGE$_2$ was administered. After 8 hours of the second dose of PGE$_2$, if reassessment did not reveal an established labour or a favourable cervix, induction was considered as failed and a caesarean section was offered. Maximum of 2 doses of PGE$_2$ (3 mg) was used.

After admitting to labour ward induction was continued with routine amniotomy followed by a oxytocin infusion. Intrapartum care and monitoring was conducted according to NICE guideline 190$^7$ and vaginal examinations performed accordingly. Successful induction of labour was the primary outcome measure. When cervical dilatation reached 8cm, induction of labour was considered as successful irrespective of the mode of delivery, as a caesarean section carried out after 8cm of dilatation was unlikely to be due to failed induction.$^{28}$ Caesarean sections performed due to fetal distress or uterine hyper-stimulation before reaching the cervical dilatation of 8cm were also not considered as failed induction and censored in the analysis. The secondary outcome measure was the time interval from amniotomy to achieve successful induction.

The distribution of continuous variables was tested for normality. In the univariate analysis Chi-square test for categorical variables and Mann-Whitney U test and unpaired Student t-test for numerical variables were used as appropriate. Multiple logistic regression analysis performed to determine usefulness of different independent variables in predicting successful induction independently. The performance of TVSCL and Bishop as tests to predict successful induction was evaluated using receiver-operating characteristic curves (ROC). Pearson correlation was used to analyse the relationship between Bishop Score, TVSCL and amniotomy – successful induction time interval. Data was analysed using statistical package for social sciences (SPSS) version 17.$^{29}$ Statistical significance was defined by $P$ values of 0.05 or less.

Ethical approval was obtained from the Ethical Review Committee, Faculty of Medicine, University of Peradeniya, Sri Lanka and informed written consent was obtained from all the participants.

**RESULTS**

A total of 428 nulliparous women were enrolled during the study period. 36 (8.4%) women who delivered by caesarean section (CS) due to uterine hyper-stimulation or fetal distress before reaching the primary outcome were not considered in the analysis in agreement with many previous authors as they were not true induction failures.

The mean age of the study population was 25.3 years (SD±3.4). The mean body mass index (BMI) and the mean maternal height were 27.5 Kg/m$^2$ (SD±3.7) and 153.6cm (SD±6.3) respectively. The majority of the population were Singhalese (79.8%) and rest of the participants were Muslims and Tamils. The mean fetal birth weight was 3200 grams (SD±30).

The distribution of Bishop score and TVSCL in the study population were approximately symmetric. The mean pre-induction Bishop score and mean TVSCL of the study population were 3.9 (SD±0.8) and 29.7mm (SD±7.1) respectively.

The distribution of continuous variables and Bishop score between successful and failed induction groups. More detailed description of induction success rates according TVSCL have been described in table 3. Independent prediction value of the variables in predicting successful labour induction was analysed by multiple logistic regression analysis (Table 4) and showed that only Bishop score and not the cervical length was a significantly independent predictor of successful induction of labour.

### Table 1 – Distribution of mode of delivery in the study sample

<table>
<thead>
<tr>
<th>Mode of delivery</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Successful induction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaginal delivery</td>
<td>241</td>
<td>61.5</td>
</tr>
<tr>
<td>Vacuum</td>
<td>18</td>
<td>4.6</td>
</tr>
<tr>
<td>Forceps</td>
<td>19</td>
<td>4.8</td>
</tr>
<tr>
<td>LSCS (Dilation≥8cm)</td>
<td>18</td>
<td>4.6</td>
</tr>
<tr>
<td>Failed induction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSCS</td>
<td>96</td>
<td>24.5</td>
</tr>
<tr>
<td>Total</td>
<td>392</td>
<td>100.0</td>
</tr>
</tbody>
</table>

LSCS = Lower segment caesarean section
Table 2 – Maternal and cervical characteristics of two groups of women who had successful induction and failed induction

<table>
<thead>
<tr>
<th>Variable</th>
<th>Successful induction (n=296)</th>
<th>Failed induction (n=96)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age, mean± SD (years)</td>
<td>24.7 ± 3.3</td>
<td>27.3 ± 2.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BMI, mean± SD (kg/m²)</td>
<td>26.8 ± 3.5</td>
<td>27.8 ± 3.8</td>
<td>0.017</td>
</tr>
<tr>
<td>Birth weight, mean± SD (g)</td>
<td>3200 ± 0.2</td>
<td>3300 ± 0.3</td>
<td>0.001</td>
</tr>
<tr>
<td>TVS cervical length, mean± SD (mm)</td>
<td>29.5 ± 6.9</td>
<td>31.4 ± 8.0</td>
<td>0.022</td>
</tr>
<tr>
<td>Bishop score, mean (95% CI)</td>
<td>4.1 (3.9 – 4.1)</td>
<td>3.7 (3.4 – 3.8)</td>
<td>&lt;0.002</td>
</tr>
</tbody>
</table>

BMI, Body mass index; TVS, transvaginal ultrasound scan
*Not significant
*Mann Whitney U test, †Unpaired t test, used.

Table 3 – Successful induction rates according to transvaginal cervical length (TVSCL) of women in the study sample

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Status of successful induction</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Successful</td>
<td>Failure</td>
</tr>
<tr>
<td>TVS Cervical Length (mm)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 – 19</td>
<td>47</td>
<td>10</td>
</tr>
<tr>
<td>20 – 24</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>25 – 29</td>
<td>86</td>
<td>12</td>
</tr>
<tr>
<td>30 – 34</td>
<td>72</td>
<td>20</td>
</tr>
<tr>
<td>≥35</td>
<td>76</td>
<td>42</td>
</tr>
<tr>
<td>Total</td>
<td>296</td>
<td>96</td>
</tr>
</tbody>
</table>

TVS=Transvaginal ultrasound scan

Table 4 – Result of logistic regression analysis of the likelihood of successful induction in terms of the selected components

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted Odds Ratio</th>
<th>95% Confidence Interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group</td>
<td>0.21</td>
<td>0.09 – 0.46</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Birth weight</td>
<td>0.35</td>
<td>0.19 – 0.62</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>0.36</td>
<td>0.15 – 0.88</td>
<td>0.03</td>
</tr>
<tr>
<td>Bishop score</td>
<td>1.99</td>
<td>1.38 – 2.88</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>TVS cervical length</td>
<td>0.88</td>
<td>0.30 – 1.03</td>
<td>0.07</td>
</tr>
</tbody>
</table>

For detailed analysis, ROC curves were constructed to evaluate the predictive value of the Bishop score and TVSCL. A new Bishop score was calculated for each individual woman replacing the cervical length obtained by digital cervical examination with TVSCL and a third ROC curve was also constructed. All three ROC curves are showed in figure 1. Performance of each assessment method in predicting successful induction is presented in table 5.

The ROC curves indicated that Bishop score ≥3 and ultrasonic cervical length ≤3.3cm were the best cut-off levels for prediction of successful induction.
Table 5 – Test characteristics of Bishop score, ultrasonographic cervical length and new Bishop score in predicting successful induction

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Sensitivity (% (95% CI,P))</th>
<th>Specificity (% (95% CI,P))</th>
<th>PPV (% (95% CI,P))</th>
<th>NPV (% (95% CI,P))</th>
<th>LR+ (% (95% CI,P))</th>
<th>LR- (% (95% CI,P))</th>
<th>AUC (% (95% CI,P))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bishop score</td>
<td>76.0 (70.7-80.8)</td>
<td>44.8 (34.6-55.3)</td>
<td>80.9 (75.8-85.4)</td>
<td>37.7 (28.8-47.3)</td>
<td>1.38 (1.1-1.7)</td>
<td>0.54 (0.4-0.7)</td>
<td>0.598 (0.548-0.647)</td>
</tr>
<tr>
<td>&gt; 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TVSCL ≤ 3.3mm</td>
<td>74.3 (69.0-79.2)</td>
<td>43.7 (33.6-54.3)</td>
<td>80.3 (75.1-84.8)</td>
<td>35.6 (27.0-44.9)</td>
<td>1.32 (1.0-1.7)</td>
<td>0.59 (0.5-0.8)</td>
<td>0.545 (0.494-0.595)</td>
</tr>
<tr>
<td>New Bishop score≥ 2</td>
<td>97.3 (94.7-98.8)</td>
<td>27.0 (18.5-37.1)</td>
<td>80.4 (76.0-84.4)</td>
<td>76.5 (58.5-89.4)</td>
<td>1.33 (1.0-1.9)</td>
<td>0.1 (0.05-0.2)</td>
<td>0.586 (0.536-0.636)</td>
</tr>
</tbody>
</table>

PPV = Positive predictive value, NPV = Negative predictive value, LR+ = Positive likelihood ratio, LR- = Negative likelihood ratio, AUC = Area under the ROC curve, TVSCL = transvaginal ultrasonographically measured cervical length

In our study 250 women had an amniotomy followed by oxytocin infusion, subsequent to the PGE₂ administration. The mean, time interval from amniotomy to successful induction (TI) was 5.23hrs (SD±2.25). Scatter graphs in figure 2 and figure 3 demonstrate the association between time interval and Bishop’s score and TVS cervical length respectively. Pearson’s correlation coefficient revealed a weak negative correlation between the Bishop score and TI (R=0.33, P=0.01) and the correlation between TVSCL and TI was not statistically significant (R=0.02, P=0.06).

DISCUSSION

Our study shows that ultrasonographic cervical length (TVCL), despite their objective character, was not an effective predictor of successful induction. The area under the curve for cervical length was not significantly different from the non-diagnostic line of the receiver-operating characteristic curve (AUC=0.545; 95% CI, 0.494-0.545; P=0.21). Our findings are in agreement with those of Chandra et al.22, Watson et al.30 and Rozenberg et al.31 who noted no statically significant predictive value of TVSCL in determining successful induction but with Bishop score. Hatfield et al.32 also failed to demonstrate the diagnostic accuracy of TVSCL in predicting successful induction in their recently published systematic review with meta-analysis.

However other authors found conflicting results. Pandis et al.33 reported a study of 240 women undergoing labour induction and showed that TVCL performed better than Bishop score to predict successful induction. Gabriel et al.28 studied 179 women underwent...
induction and reported that TVSCL had a better predictive value for successful induction compared to Bishop score.

The contradictory results of these studies are likely to be explained by many reasons. There were significant heterogeneity among studies with regard to demographic characteristics, parity, indication of induction, gestational age, method of induction and pre-induction cervical favourability status. In several instances, different methods of inductions were used in the same study.

Unlike the other studies, we specifically focus on a population of nulliparous women beyond 40 weeks of gestation and who had Bishop score of 5 or less. Therefore strict comparison of their results with our results is not rational.

We constructed a ROC curve for a newly calculated Bishop score which derived incorporating the TVSCL instead of cervical length obtained by vaginal examination. Although AUC of this new score is not significant, the best cut-off value demonstrates a higher sensitivity than traditional Bishop score (97.3 vs. 76). This finding exhibit an indirect evidence of the importance of other individual parameters of the Bishops score than cervical length alone, in predicting successful induction and emphasizes the value of incorporating multiple ultrasonic parameters like cervical funnelling, posterior cervical angle and fetal head-perineum distance which can substitute the parameters of Bishop score.

Although test characteristics of Bishop score shows some usefulness in determining successful induction in our study population, its positive likelihood ratio is only 1.38 and indicates its poor discriminatory value. The positive likelihood ratio of the Bishop score found to vary 1.3-3.0 throughout the literature. Bishop’s original study was mainly focused on multiparous women and there was no statistical analysis to confirm the validity of the score as a diagnostic test. It is widely considered that a diagnostic test should have a positive LR of 5 or greater or a negative LR of 0.2 or less to be clinically useful\textsuperscript{14}, both methods were clearly suboptimal when measured against this standard. It could be concluded therefore that Bishop score is also ineffective as a tool to predict success of induction of labour in this group of pregnant women. This does not mean that the Bishop score should be abandoned, as it may be useful in determining whether the cervix is ripe or whether further doses of prostaglandins are required to achieve ripeness. However, when used alone to determine the likely outcome of induction of labour, the Bishop score...
does not achieve current standards required of an effective diagnostic test in this group of women.

Finally, in agreement with several previous studies, our results were also unable to demonstrate any significant value of TVSCCL in predicting amnionotomy to successful induction time interval. It is well known that, in addition to cervical length, many other confounding factors related to uterine behaviour, maternal passage and the fetes have significant influence on determining the duration of the first stage of labour.

LIMITATIONS
In our study we used prostaglandin vaginal tablet as the only ripening agent and only maximum of two doses used. Different results may have been received from other preparations and more doses.

CONCLUSION AND RECOMMENDATIONS
Transvaginal ultrasonographic measured cervical length does not demonstrate a diagnostic accuracy in predicting labour induction outcomes; successful induction or time interval to successful induction compared to digital cervical assessment in terms of Bishop score in nulliparous pregnancies beyond 40 weeks with unfavourable cervix.

Future studies should be focused on recognition the value of other multiple potential ultrasonic parameters like cervical wedging, posterior cervical lip and fetal head perineum distance which can substitute the parameter of Bishop score and to integrate and validate them in terms of predicting successful induction.

REFERENCES
22. Chandra S, Crane JM, Hutchens Distributed, Young DC. Transvaginal ultrasound and digital examination