Matijević R., Grgić O., Vasilj O. (2006) ***Is sonographic assessment of cervical length better than digital examination in screening for preterm delivery in a low-risk population?*** **Acta obstetricia et gynecologica Scandinavica, 85 (11). pp. 1342-1347. ISSN 0001-6349**

http://www.informaworld.com/smpp/title~db=all~content=t716100748

http://dx.doi.org/10.1080/00016340600935722

Is sonographic assessment of cervical length better than digital examination in screening for preterm delivery in low-risk population?

Ratko Matijevic

Ozren Grgic

Oliver Vasilj

Department of Obstetrics and Gynecology, School of Medicine, Zagreb University, Sveti Duh Hospital, Zagreb, Croatia

Correspondence author address:

Dr Ozren Grgic

Sveti Duh Hospital

University Department of Obstetrics and Gynecology

Sveti Duh 64

10000 ZAGREB

Croatia

Tel: +385 1 37 123 17

Fax: +385 1 37 455 34

E-mail: ozren.grgic@gmail.com

Running headline: Screening tests for prematurity in low-risk population

Is sonographic assessment of cervical length better than digital examination in screening for preterm delivery in low-risk population?

Abstract

Background: This randomized controlled trial compared the diagnostic accuracy of the sonographic assessment of cervical length and clinical digital examination of the cervix in the second trimester regarding the prediction of preterm delivery in low-risk population.

Methods: Totally, 282 unselected, asymptomatic, women with singleton pregnancy randomly underwent sonographic cervical length measurement (study group, n=138) or clinical digital examination (control group, n=144) in the second trimester. In the study group cervical length $\leq 5^{th}$ percentile ($\leq$ 24mm) for our population was defined as shortened. In the control group, Bishop score $\geq 95^{th}$ percentile ($\geq$ 4) for our population was defined as high. Primary outcome measure included the diagnostic accuracy of both tests regarding the prediction of preterm delivery (< 37 weeks).

Results: Shortened cervical length was found in 6/138 (4.3%) whereas the high Bishop score was found in 17/144 (11.8%) (p = 0.038, Fisher exact test). The incidence of preterm delivery was 5.7 % (16/282). Regarding the prediction of preterm delivery, shortened cervical length and high Bishop score had sensitivity 57.1 % vs. 33.3 % and positive predictive value 66.7 % vs. 17.6 %. Shortened cervical length in comparison with high Bishop score had 12-fold higher positive likelihood ratio for preterm delivery in low-risk population (37.4; 95%CI [8.2-170.7] vs. 3.2; 95%CI [1.1-9.2]).

Conclusion: Sonographic assessment of cervical length has better diagnostic accuracy in the prediction of preterm delivery compared to digital examination in low-risk population.

Keywords: low-risk population; transvaginal cervical sonography; digital examination; Bishop score; second trimester; preterm delivery

Abbreviation and units

CI – confidence interval

CL – cervical length

DE – digital examination

PTD – preterm delivery

LR – likelihood ratio

NICU – neonatal intensive care unit

NPV – negative predictive value

PPV – positive predictive value

RCT - randomized controlled trial

TVS – transvaginal sonography

Introduction

The prediction of preterm delivery (PTD) in asymptomatic low-risk population relies on different clinical, ultrasound based and biochemical screening tests. All of them are characterized by low sensitivity and low positive predictive value (PPV) [1-5]. As a consequence of such improper selection, some low-risk women may be unnecessarily monitored, hospitalized and even exposed to potentially dangerous tocolytic agents. On the other hand, pregnant women at risk for PTD may remain unrecognized and consequently have a late institution of corticosteroid prophylaxis and tocolytics or a delayed transfer to tertiary center, resulting in increased perinatal mortality and morbidity [4, 5].

Clinical digital examination (DE) of the cervix is still routinely used in clinical practice for assessment of low and high-risk pregnant women [6-11]. This examination is simple, cheap, and it is believed to be useful in the assessment of the cervical consistency, shape, station and dilatation, in order to recognize women at risk for PTD [7, 9]. Despite being abandoned in some western countries [12], in most parts of the world DE is still performed at each clinical appointment [6-11].

Transvaginal sonography (TVS) of the uterine cervix has been suggested as a method for the quantitative and qualitative cervical evaluation in low-risk pregnancy. Among the different sonographic parameters used, majority of investigators have focused on cervical length (CL) measurement as a well-accepted and standardized method of cervical assessment [1-3, 13]. Additionally, CL measurement has better accuracy compared to the other sonographic parameters (i.e. funneling) regarding the prediction of PTD in low-risk population [14 –15].

The advantage of TVS of the cervix over the DE is related to the possibility of assessment of supravaginal portion and internal cervical axis; both being impossible to evaluate digitally [16]. Sonographic assessment allows quantitative and qualitative measurement of the cervix reducing the inter and intra-observer variations [17-18]. Presently, the comparison of TVS

and DE was mostly performed in high-risk population [7, 19-21], and the results of randomized controlled trials (RCT) of their usefulness in low-risk population are still missing. Therefore, in this RCT we compared the diagnostic accuracy of CL measurement assessed by TVS and clinical DE in the second trimester, regarding the prediction of PTD in asymptomatic low-risk population.

Materials and methods

Study design and subjects

This RCT was a part of a research project designed to assess the efficacy of different diagnostic methods as potential screening tests for late spontaneous miscarriage and preterm delivery (Croatian Ministry of Science, Project No. 0129111). The target population was: asymptomatic, nulliparous women with uncomplicated singleton pregnancy attending sonographic or clinical appointment between 16 and 23 completed weeks. At that gestational age, all women were routinely booked for anomaly scans and those satisfying our inclusion criteria were approached. The exclusion criteria were: 1.Clinicaly and laboratory suspected infection (White cell count $\geq$ 14 x 10^{12}/l, C reactive protein $\geq$ 10 mg/l) at enrollment, 2. Evidence of cervical and vaginal infection on swabs, if taken, 3. History of surgical procedure on the cervix, 4. Developmental malformations of the Müllerian ducts, 5. Cervical cerclage before enrollment, 6. Other medical conditions that are risk factors for preterm delivery, 7. Major congenital fetal anomalies, and 8. Intrauterine fetal death.

Local and national ethical committees approved the study protocol and all women included in the study gave their informed written consent.

Randomization

Computer-generated random number tables were used for random assignment. Group assignments were placed in sealed, opaque, alternate numbered envelopes. Pregnant women were randomly assigned into two groups: sonographic assessment of CL (study group) or DE of the cervix (control group).

The results obtained in the study were known to the principal investigators only and the course of the pregnancy was unaltered by the participation in the study.

Collection of data and sonography

Transvaginal sonography was performed with a 7-MHz transducer with an angle of 160^{o} (model 6117, Aloka 5500, ALOKA CO. LTD, Japan). Each TVS was performed according to the protocol described by Owen et al [21]. This method has a mean interobserver variation of 4–10% and a mean intraobserver variation of 5% [17-18].

An adequate image was defined as presence of external cervical axis, endocervical canal and internal cervical axis, which is the cranial end of the cervical canal adherent to the intrauterine cavity. Cervical length was measured with electronic calipers as linear distance between the external axis and the functional internal axis along a closed endocervical canal, and the length below the 5^{th} percentile for our population was defined as shortened. The sonographic assessment of the cervix was carried out over a period of three minutes to detect spontaneous dynamic changes. At least three measurements were made and the lowest value was used for further calculations.

In the control group a DE was performed to assess the components of the Bishop score, a composite measure that assigns a score of 0 to 3 points to each of four features of the cervix: length, dilatation, position, consistency, and station of the presenting part [22]. Bishop score above the 95^{th} percentile for our population was defined as high.

Data analysis

As a part of the study design, a sample size calculation was performed (Sample size calculator, MaCorr Inc., Toronto, Ontario, Canada). The calculation was designed to detect at least 10 % difference between the sensitivity of CL measurement and clinical DE, regarding the prediction of PTD in low risk population. Confidence interval (CI) was determined using the confidence level of 95%. The program has calculated that CI is 5.9 %. Using confidence level of 95% and CI of 5.9 % the program has calculated that for this population the sample size needs to be 276 participants (138 in each group).

Data were analyzed using SPSS version 11.0 (SPSS Inc., Chicago IL, USA). Cervical length (continuous variable) and Bishop score (composite measure of four features of the cervix: length, dilatation, position, consistency, and station of the presenting part) were the variables used to predict outcomes.

Primary outcome measure included sensitivity, specificity, PPV and negative predictive value (NPV) with 95% CI for shortened CL ($\leq$ 24mm) and high Bishop score ($\geq$ 4) regarding the prediction of PTD (< 37 weeks). Percentiles for the CL and Bishop score were analyzed by the use of chi-square tests. We calculated likelihood ratios (LR) for positive results and 95 % CI comparing subjects at or below 5^{th} percentile with those above the 5^{th} percentile in the study group whereas in the control group we compared subjects at or above the 95^{th} percentile with those below 95^{th} percentile.

Secondary outcome measures included: gestational age at delivery, incidence of early PTD (< 34 weeks), iv tocolysis and corticosteroid administration, birth weight, perinatal mortality rate and perinatal morbidity assessed by Apgar score $\leq$ 7 at 5 minutes and admission into the neonatal intensive care unit (NICU).

Fisher exact test was used to evaluate categorical data where appropriate. A Student t-test was used to evaluate continuous variables. Two-tailed p values were reported throughout and statistical significance was defined as $p < 0.05$.

Results

During the study period (October 2002 to April 2005) 369 women were approached. The flow diagram of enrollment and exclusions is presented in Figure I. Final results were based on 282 pregnant women.

The incidence of PTD < 37 weeks was 5.7 % (16/282) (study group 5.1 (7/138) vs. control group 6.2 (9/144), p= 0.865, Fisher exact test). Maternal age (mean ± standard deviation) at randomization was 28.1 ± 5.3 years (study group 28.1 ± 4.7 vs. control group 28.1 ± 5.8, p=0.984, Student t-test) and gestational age at randomization was 19.9 ± 2.1 weeks (study group 19.9 ± 2 vs. control group 19.9 ± 2.2, p=0.722, Student t-test). Cervical length was 39.5 ± 8.5 mm (range 10 – 65 mm), and the median of Bishop score was 1 (interquartile range 0 – 2). The diagnostic accuracy of shortened CL (≤ 24 mm) and high Bishop score (≥ 4) regarding the prediction of PTD < 37 weeks is presented in Table I.

Regarding the prediction of PTD < 37 weeks, shortened CL and high Bishop score had sensitivity 57.1 % vs. 33.3 % and PPV 66.7 % vs. 17.6 %. The specificity and NPV were similar. Shortened CL measurement in comparison with high Bishop score had 12 fold higher positive LR for PTD < 37 weeks in low-risk population (37.4; 95%CI [8.2-170.7] vs. 3.2; 95%CI [1.1-9.2]).

The comparison of secondary outcome measures between the study and control group is presented in Table II. There was one case of perinatal death in the study group. The woman had premature preterm rupture of membranes at 25th week, and delivered in 26th week, male fetus weighting 950 grams. The newborn died 24 hours later because of the complications caused by severe respiratory distress. There was no statistically significant difference between the study and control group in all assessed parameters.

Discussion

Preterm delivery is one of the biggest problems in modern obstetrics [4, 23]. Despite the fact that perinatal mortality is relatively low after 32 completed gestational weeks, so called "marginally" PTD (35 - 37 weeks) remains responsible for a relatively high rate of short and long-term neonatal morbidity, and it is still important problem in clinical practice [23].

The target population in our RCT was low-risk, asymptomatic primgravidae with singleton pregnancy. Presently available screening tests for PTD in this population have low sensitivity and low PPV. Despite the fact that prevalence of PTD in this population is low, those women are important as they account for almost half of all pregnancies. This justifies the need for screening but also makes the design of screening protocols very difficult. In order to get acceptable specificity the cut off values of measurable parameters have to be set below 5^{th} or above 95^{th} percentile [1-3,13].

Among the different sonographic screening tests, a CL measurement assessed by TVS has the greatest diagnostic accuracy regarding the prediction of PTD in low-risk population, being superior compared to the other presently available ultrasound based screening tests (i.e. funneling) [14-15]. The distribution of CL measurements in our study (39.5 $\pm$ 8.5 mm) was similar to the results published in the literature, i.e. Heath et al 38 mm [1]; Iams et al 35.2 $\pm$ 8.3 mm (at 24 weeks) [2]; Taipale et al 40.7 $\pm$ 7 mm [3]; and Hassan et al 37.5 $\pm$ 6.6 mm [13]. Due to the relatively low number of participants included in this trial, the diagnostic accuracy of our results regarding CL measurement cannot be compared with other trials. However, this was not our goal, as we wanted to compare the diagnostic accuracy between CL measurement and DE in the low-risk population.

Cervical assessment in pregnancy by DE has been the traditional method for prediction of PTD. According to the literature a routine DE of the cervix has not resulted in a decreased incidence of PTD mostly due to poor PPV. This consequently may influence the number of

unnecessary hospital admissions or tocolytic use [12]. Among the different parameters assessed by DE majority of investigators have used the Bishop score, a composite measure of four features of the cervix and station of the presenting part [22, 24].

In the high-risk population, DE of the cervix had very limited value with very poor sensitivity and PPV [19-21]. Despite all of that, DE of the cervix is still systematically and widely used in everyday clinical practice for routine assessment of low-risk and high-risk pregnant women [6-11]. The only advantage of this screening test is in its simplicity, availability, price and no equipment requirements.

The comparison of diagnostic accuracy between CL measurement and DE in our study with the results of Andersen et al [25], found similar sensitivity regarding the prediction of PTD for CL measurement (57%, 95%CI [20 - 88] vs. 47%, 95%CI [4 - 72]) and similar sensitivity for DE (33%, 95%CI [9 - 69] vs. 36%, 95%CI [14% to 64%]).

Most of investigators found that objectivity and reliability of CL measurement assessed by TVS justifies its use in screening for PTD in high-risk population. Some trials have shown that sonographic measurement of CL in low-risk population might help to identify the women at risk for PTD [1-3, 13]. Therefore, we compared those two methods in low-risk population. The only problem is the lack of the standardized diagnostic criteria for cervical incompetence in this population, as presently we do not know which criterion performs the best. At the moment the evidence is in favor of measurements of the length of closed part of the cervix, and therefore we used it in our trial [1-3, 13].

The results of our RCT confirm that DE of the cervix has a lower sensitivity and lower PPV in comparison to the CL measurement regarding the prediction of PTD in low-risk population. Shortened CL measurement in comparison with high Bishop score had 12-fold higher positive LR for PTD < 37 weeks in low-risk population. This difference is mostly due to the fact that

DE of the cervix is very subjective measurement (interobserver variability of > 50%) [26], and it is not accurate for evaluation of the internal cervical axis as the whole upper part of the cervix is not assessable by this method [16]. As well as that, DE is nonspecific examination, as up to 15% of primiparous women, who have delivered at term, had positive findings on DE in the late second trimester [27].

We are aware that our sample size is relatively low compared to the other studies assessing shortened CL regarding the prediction of PTD in the low-risk population. However, this is a first RTC that compared the diagnostic accuracy of the sonographic assessment of CL and clinical DE of the cervix in the second trimester regarding the prediction of PTD in low-risk population. Our results show that CL measurement compared to DE has better diagnostic accuracy in screening for PTD in low-risk population. The use of CL measurement could improve the selection of woman at risk and reduce perinatal mortality and morbidity by enabling the appropriate administration of corticosteroids and transfer to tertiary center. Additionally, it may reduce the use of tocolytic agents, and the frequency of the side effects that are related with them [28].

References

1. Heath VC, Southall TR, Souka AP, Elisseou A, Nicolaides KH. Cervical length at 23 weeks of gestation: prediction of spontaneous preterm delivery. Ultrasound Obstet Gynecol 1998; 12: 312-7.

2. Iams JD, Goldenberg RL, Meis PJ, Mercer BM, Moawad A, Das A et al. The length of the cervix and the risk of spontaneous premature delivery. N Engl J Med 1996; 334: 567-72.

3. Taipale P, Hiilesmaa V. Sonographic measurement of uterine cervix at 18–22 weeks' gestation and the risk of preterm delivery. Obstet Gynecol 1998; 92: 902-7.

4. Mercer BM, Goldenberg RL, Das A, Moawad AH, Iams JD. The preterm prediction study: a clinical risk assessment system. Am J Obstet Gynecol 1996; 174(6): 1885-95.

5. Iams JD, Goldenberg RL, Mercer BM, Moawad AH, Meis PJ, Das AF et al. Can low-risk women destined for spontaneous preterm birth be identified? Am J Obstet Gynecol 2001; 184: 652-5.

6. Jackson GM, Ludmir J, Bader TJ. The accuracy of digital examination and ultrasound in the evaluation of cervical length. Obstet Gynecol 1992; 79(2): 214-8.

7. Volumenie JL, Luton D, De Spirlet M, Sibony O, Blot P, Oury JF. Ultrasonographic cervical length measurement is not a better predictor of preterm delivery than digital examination in a population of patients with idiopathic preterm labor. Eur J Obstet Gynecol Reprod Biol 2004; 117(1): 33-7.

8. Sergent F, Marpeau L. For transvaginal digital examination in obstetrics. Gynecol Obstet Fertil 2005; 33(1-2): 65-8.

9. Rozenberg P. Intravaginal ultrasonography of the uterine cervix: hope in the campaign against premature births. J Radiol 1999; 80(5): 421-9.

10. Zlatnik FJ, Yankowitz J, Whitham J, Burmeister LF. Vaginal ultrasound as an adjunct to cervical digital examination in women at risk of early delivery. Gynecol Obstet Invest 2001; 51:12-6.

11. Matijevic R, Grgic O. Clinical examination and transvaginal sonograpy in mid trimester as potential screening tests for preterm labor. Preliminary results on low risk population. J Matern Fetal Neonatal Med 2004;16 Suppl 1: 48.

12. Buekens P, Alexander S, Boutsen M, Blondel B, Kaminski M, Reid M. Randomised controlled trial of routine cervical examinations in pregnancy. Lancet 1994; 334: 841-4.

13. Hassan SS, Romero R, Berry SM, Dang K, Blackwell SC, Treadwell MC et al. Patients with an ultrasonographic cervical length 5 or=15 mm have nearly a 50% risk of early spontaneous preterm delivery. Am J Obstet Gynecol 2000; 182: 1458-67.

14. Guzman ER, Walters C, Ananth CV, O'Reilly-Green C, Benito CV, Palermo A et al. A comparison of sonographic cervical parameters in predicting spontaneous preterm birth in high-risk singleton gestations. Ultrasound Obstet Gynecol 2001; 18: 204-10.

15. To MS, Skentou C, Liao AW, Cacho A, Nicolaides KH. Cervical length and funneling at 23 weeks of gestation in the prediction of spontaneous early preterm delivery. Ultrasound Obstet Gynecol 2001; 18(3): 200–3.

16. Michaels WH, Montgomery C, Karo J, Temple J, Ager JA, Olson J. Ultrasound differentiation of the competent from the incompetent cervix: prevention of preterm delivery. Am J Obstet Gynecol 1986; 154: 537-46.

17. Burger M, Weber-Rossler T, Willmann M. Measurement of the pregnant cervix by transvaginal sonography: an interobserver study and new standards to improve the interobserver variability. Ultrasound Obstet Gynecol 1997; 9: 188-93.

18. Zorzoli A, Soliani A, Perra M, Caravelli E, Galimberti A, Nicolini U. Cervical changes throughout pregnancy as assessed by transvaginal sonography. Obstet Gynecol 1994; 84: 960-4.

19. Berghella V, Tolosa JE, Kuhlman K, Weiner S, Bolognese RJ, Wapner RJ. Cervical ultrasonography compared with manual examination as a predictor of preterm delivery. Am J Obstet Gynecol. 1997; 177(4): 723-30.

20. Gomez R, Galasso M, Romero R, Mazor M, Sorokin Y, Goncalves L et al. Ultrasonographic examination of the uterine cervix is better than cervical digital examination as a predictor of the likelihood of premature delivery in patients with preterm labor and intact membranes. Am J Obstet Gynecol. 1994; 171(4): 956-64.

21. Owen J, Yost N, Berghella V, Thom E, Swain M, Dildy GA et al. Mid-trimester endovaginal sonography in women at high risk for spontaneous preterm birth. JAMA 2001; 286:1340-8.

22. Bishop EH. Pelvic scoring for elective induction. Obstet Gynecol 1964; 24: 266-8.

23. Seubert DE, Stetzer BP, Wolfe HM, Treadwell MC. Delivery of the marginally preterm infant: what are the minor morbidities? Am J Obstet Gynecol 1999; 181: 1087-91.

24. Blondel B, Le Coutour X, Kaminski M, Chavigny C, Breart G, Sureau C. Prediction of preterm delivery: is it substantially improved by routine vaginal examinations? Am J Obstet Gynecol 1990; 162: 1042-8.

25. Andersen HF, Nugent CE, Wanty SD, Hayashi RH. Prediction of risk for preterm delivery by ultrasonographic measurement of cervical length. Am J Obstet Gynecol. 1990; 163: 859-67.

26. Holcomb VL, SmeltzerJS. Cervical effacement: variation in belief among clinicians. Obstet Gynecol 1991; 78: 43-5.

27. Floyd WS. Cervical dilatation in the mid-trimester of pregnancy. Obstet Gynecol 1961; 18:380-1.

28. Anotayanonth S, Subhedar NV, Garner P, Neilson JP, Harigopal S. Betamimetics for inhibiting preterm labour. Cochrane Database Syst Rev. 2004;18;(4):CD004352.

Acknowledgment

This trial was supported by Croatian ministry of science, education and sport, grant No 0129111

Table I

Diagnostic indexes and predictive values of cervical length measurement (study group) and Bishop score (control group) regarding the prediction of preterm delivery

	Study group (CL measurement)	Control group (BS measurement)
n	138	144
Cut off value (percentile)	$\leq$ 24 mm ($\leq 5^{th}$ percentile)	$\geq$ 4 ($\geq 95^{th}$ percentile)
Positive test result (%, n)	4.3 % (6/138)	11.8 % (17/144)
Sensitivity (%, [95% CI], n)	57.1 [20.2-88.2], 4/7	33.3 [9 - 69.1], 3/9
Specificity (%, [95% CI], n)	98.4 [94 - 99.7], 129/131	89.7 [82.9 - 94], 121/135
PPV (%, [95% CI], n)	66.7 [24.1 - 94], 4/6	17.6 [4.7 - 44.2], 3/17
NPV (%, [95% CI, n)	97.7 [93 - 99.4], 129/132	95.3 [89.6 - 98], 121/127
LR+ [95%CI]	37.4 [8.2 - 170.7]	3.2 [1.1 - 9.2]

Abbreviations: CL – cervical length, BS – Bishop Score, n – number, CI – confidence interval, PPV – positive predictive value, NPV – negative predictive value, LR+ - likelihood ratio for positive results

Table II

Primary and secondary outcomes measure

	Study group (CL measurement)	Control group (BS measurement)	Statistical significance	
			p	RR [95% CI]
Primary outcome measure				
Delivery < 37 weeks	7	9	0.671 (1)	0.81 [0.31 – 2.12]
Secondary outcome measures				
Gestation at delivery (CW – mean +/- SD)	38.8 +/- 1.8	38.5 +/- 2.4	0.214 (2)	NA
Delivery ≤ 34 weeks	3	4	0.962 (1)	0.78 [0.18 – 3.43]
IV tocolysis and CA	3	10	0.104 (1)	0.31 [0.08 – 1.11]
Perinatal mortality	1	0	0.986 (1)	NA
Birth weight (grams – mean +/- SD)	3273 +/- 454	3243 +/- 626	0.642 (2)	NA
Apgar ≤ 7 after 5 min	2	3	0.712 (1)	0.69 [0.12 – 4.11]
NICU admission	1	2	0.970 (1)	0.52 [0.04 – 5.69]

Abbreviations: CL – cervical length, BS – Bishop Score, RR – risk ratio, CI – confidence interval, CW – completed weeks, SD – standard deviation, NA – not applicable, IV – intravenous, CA – corticosteroid administration, NICU – neonatal intensive care unit

(1) Fisher exact test

(2) Student t-test

Figure I

Consort flowchart of enrolment