Ultrasound Obstet Gynecol 2010; **35**: 54–64
Published online 15 December 2009 in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/uog.7457

Transvaginal cervical length measurement for prediction of preterm birth in women with threatened preterm labor: a meta-analysis

A. SOTIRIADIS*†, S. PAPATHEODOROU*, A. KAVVADIAS* and G. MAKRYDIMAS*

**Department of Obstetrics and Gynecology, University Hospital of Ioannina, Ioannina and †Iatriki Emvryou, Thessaloniki, Greece*

KEYWORDS: cervical length; preterm birth; ultrasound

ABSTRACT

Objectives *To integrate data on the performance of cervical length measurement for the prediction of preterm birth in symptomatic women.*

Methods *MEDLINE, SCOPUS and manual searches for studies with transvaginal ultrasound measurement of the cervical length in symptomatic women were carried out. Random effects models were used for data integration, and pooled test estimates of sensitivity, specificity, and positive and negative likelihood ratios (LR+ and LR−) were calculated along with their 95% CIs.*

Results *Twenty-eight studies fulfilled the selection criteria. For birth within 1 week from presentation, the pooled sensitivity, specificity, LR+ and LR− of cervical length < 15 mm were 59.9% (95% CI, 52.7–66.8%), 90.5% (95% CI, 89.0–91.9%), 5.71 (95% CI, 3.77–8.65) and 0.51 (95% CI, 0.33–0.80), respectively. The same estimates for studies with presentation at or before 34 + 0 weeks were 71.0% (95% CI, 60.6–79.9%), 89.8% (95% CI, 87.4–91.9%), 5.19 (95% CI, 2.29–11.74) and 0.38 (95% CI, 0.11–1.34), respectively. For prediction of birth before 34 weeks, the pooled sensitivity, specificity, LR+ and LR− of cervical length < 15 mm were 46.2% (95% CI, 34.8–57.8%), 93.7% (95% CI, 90.7–96.0%), 4.31 (95% CI, 2.73–6.82) and 0.63 (95% CI, 0.38–1.04), respectively. There was considerable heterogeneity across studies in most estimates.*

Conclusions *Measurement of cervical length in symptomatic women can detect a significant proportion of those who will deliver within 1 week and help to rationalize their management. The considerable heterogeneity across studies may be indicative of methodological flaws, which either were not reported at all or were underreported. Copyright © 2009 ISUOG. Published by John Wiley & Sons, Ltd.*

INTRODUCTION

Very preterm birth, usually defined as birth before 32 gestational weeks[1], is a leading cause of neonatal morbidity and mortality. Although assisted ventilation and the introduction of surfactant therapy improved pediatric outcome in the 1990s, a baby born at 24 weeks' gestation still has only about a 50% chance of survival, and approximately more than one quarter of the survivors will have severe neurodevelopmental handicap[2,3]. The rates of perinatal death and severe handicap decline to 8% and 10% for babies born at 30 weeks and further reduce to 0.4% and 5% for those born at 34 weeks, respectively[4].

Preterm labor is one of the most common reasons for prenatal admission to hospital, accounting for up to one third of such cases in some reports[5,6]. However, the diagnosis of true preterm labor remains unreliable, resulting in significant over-treatment, since fewer than 10% of these women will eventually deliver within 1 week or before 35 weeks, irrespective of treatment[7]. Apart from maternal risks, hospitalization for undelivered preterm labor has a significant financial cost.

Prediction and prevention of preterm birth have come a long way in asymptomatic women; measurement of the length of the uterine cervix can achieve a 55% sensitivity for a false positive rate of 10%[8] and administration of progesterone in women with a short (< 15 mm) cervix can decrease the likelihood of preterm birth at less than 34 weeks' gestation by 45%[9].

Cervical length measurement has also been applied in symptomatic women, and different studies report different rates for sensitivity and specificity. Moreover, many of these studies included limited numbers of actual preterm births, resulting in wide confidence intervals. In this study we systematically reviewed the literature on cervical-length measurement for the prediction of preterm birth in symptomatic women, and integrated data using meta-analytic techniques in order to acquire precise estimates

Correspondence to: Dr A. Sotiriadis, 10 P. Mela Street, 54622 Thessaloniki, Greece (e-mail: asotir@gmail.com)

Accepted: 18 September 2009

for the predictive performance of the test and explore potential causes for inconsistency across the studies.

METHODS

Sources

Two investigators (S. P. and A. S.) searched the literature (last update May 2009) for studies on the prediction of preterm birth in women with symptoms of preterm labor, using ultrasonographic measurement of the length of the uterine cervix. *MEDLINE* and *SCOPUS* searches (1960–May 2009) used combinations of the terms 'preterm birth' OR 'preterm labor' AND 'cervi*' or 'ultrasound' or 'ultrasonography'. These searches were complemented by perusal of the references of the retrieved articles and additional automated searches using the 'search for related articles' *PubMed* function. All studies were carefully compared to ensure avoidance of duplicate or overlapping samples. In case of overlap, the study with the largest number of preterm births was included.

Study selection

For our analysis we considered prospective or case-control studies evaluating transvaginal ultrasonographic assessment of cervical length for the prediction of preterm birth in women with a singleton pregnancy and intact membranes, presenting with symptoms of preterm labor. Both studies with or without tocolysis and/or prophylactic steroid administration were included; English and non-English language studies were considered. Abstracts were considered only if they provided adequate information.

Studies with multiple pregnancies or premature rupture of membranes were excluded, unless it was certain that such cases represented less than 20% of the total. Included studies were perused for cases of iatrogenic (induced) preterm birth, and every effort was made to exclude such cases from the analysis. Recorded information was independently entered into pre-defined forms by two researchers. Discrepancies between the two researchers were evaluated by a third researcher and consensus was reached for all data.

The primary outcomes of the meta-analysis were birth within 48 h from presentation, birth within 1 week from presentation and birth before 34 weeks' gestation, with secondary outcomes being birth before 32, 35, 36 or 37 weeks' gestation.

Statistical methods

The extracted data were used to construct 2×2 tables (true positive, false negative, false positive, true negative) for the different cervical lengths and gestational ages at birth. The analysis was performed based on the random effects model (der Simonian-Laird), which, in the presence of between-study heterogeneity, is more conservative, providing wider confidence intervals. When there is no observed between-study heterogeneity, the fixed and random effects coincide[10]. The predictive performance of cervical length measurement was assessed by means of weighted sensitivity and specificity values, positive and negative likelihood ratios (LR+ and LR−) and summary receiver–operating characteristics curve (SROC) analysis.

LR values of 1 suggest no discriminating ability at all, whereas the diagnostic performance is better the higher the LR+ and the lower the LR−. LR− < 0.2 is thought to provide strong diagnostic evidence in rule-out situations[11]. SROC curve analysis accounts for the interdependence between weighted sensitivity and specificity, and uses a regression method to fit a curve that best describes the trade-off of the two estimates across studies.

Heterogeneity

Between-studies heterogeneity was assessed using the I^2 statistic, which describes the percentage of variation that is due to heterogeneity rather than chance. This was preferred over the Q-statistic, as the latter may have low power in assessing heterogeneity when the number of studies is small. Although the quantification of heterogeneity is rather arbitrary, the adjectives low, moderate, and high (inconsistency) are tentatively assigned to I^2 values of 25, 50 and 75%[12].

Methodological quality and sensitivity analysis

Adequate description of test methodology and exclusion criteria was used for quality assessment in all studies. More specifically, scan methodology was considered adequate when performed as described by Heath *et al.*[13] and Andersen *et al.*[14], otherwise inadequate or not stated. Briefly, the appropriate ultrasound methodology for transvaginal measurement of cervical length involves emptying the urinary bladder, placement of the probe in the anterior fornix, acquisition of a midsagittal view of the cervix and the endocervical mucosa and measurement of the distance between the triangular area of echodensity at the external os and the V-shaped notch at the internal os. In case of heterogeneity, sensitivity analyses were performed using only studies of the highest quality. A further sensitivity analysis was done with studies in which *all* patients were recruited at or before $34 + 0$ weeks. Strict definition of preterm labor (e.g. using tocodynamometry) was not used as a quality criterion because (1) it still remains a subjective diagnosis and, (2) it could only influence the prevalence of true preterm birth, which however does not affect sensitivity and specificity.

RESULTS

The initial search identified 530 potentially relevant studies; 138 of them were retrieved for thorough

evaluation and 28 were finally eligible for synthesis with meta-analytic techniques. Sixteen of them included ultrasound examination of the cervix only[7,15–29] and 12 included both ultrasound examination and biochemical testing[30–41]. Three of these studies met the eligibility criteria but could not be used for data synthesis, either because their outcome was not included in the design of the meta-analysis[38], or because there were fewer than two studies per outcome and therefore the data could not be integrated[24,32].

The MOOSE flowchart for the meta-analysis is presented in Figure 1 and a description of the eligible studies is presented in Table 1, while Table 2 shows all the studies meeting the inclusion criteria, together with their integrated results (when appropriate) for each analyzed cut-off for preterm birth.

Birth within 48 hours from presentation

A cut-off of 15 mm for cervical length (three studies, 1266 women, rate of preterm birth within 48 h 7.1%) predicted 71.1% (95% CI, 59.5–80.9%) of preterm births at < 48 h with a specificity of 86.6% (95% CI, 84.6–88.5%)[7,16,23]. Both sensitivity and specificity showed considerable heterogeneity across the studies, despite the theoretical similarity in cervical length methodology and the lack of obvious disparities in the selection and exclusion criteria. The pooled LR+ and LR− estimates were 5.92 (95% CI, 4.91–7.13) and 0.35 (95% CI, 0.15–0.82), respectively. The performance of different cut-offs for the prediction of birth within 48 h is shown in Table 2.

Birth within 7 days from presentation

Cervical length < 15 mm (six studies, 1781 women, 197 women with preterm birth within a week (11.1%)) could predict 59.9% (95% CI, 52.7–66.8%) of preterm births within a week, with a pooled specificity of 90.5% (95% CI, 89.0–91.9%)[7,18–20,23,33]. Again, there was profound heterogeneity across studies (12 had a heterogeneity of $I^2 > 80\%$). The pooled LR+ and LR− estimates were 5.71 (95% CI, 3.77–8.65) and 0.51 (95% CI, 0.33–0.80), respectively (Figure 2).

Cervical length < 20 mm (four studies, 1263 women, 118 preterm births within 7 days (9.3%)) could predict 75.4% (95% CI, 66.6–82.9%) of preterm births within one week, with a specificity of 79.6% (95% CI, 77.1–81.9%)[7,19,30,33]. The pooled LR+ and LR− estimates were 3.74 (95% CI, 2.77–5.05) and 0.33 (95% CI, 0.15–0.73), respectively (Figure 2).

On raising the cervical length cut-off to 25 mm (four studies, 856 women, 83 cases of preterm birth within 7 days (9.7%)), the sensitivity became 78.3% (95% CI, 67.9–86.6%), with a specificity of 70.8% (95% CI, 67.4–74.0%)[15,18,19,33]. The pooled LR+ and LR− estimates were 2.81 (95% CI, 2.01–3.93) and 0.36 (95% CI, 0.24–0.54), respectively (Figure 2).

The composite SROC curve (Figure 3) presents the pairs of sensitivity and specificity for all four tested cervical length cut-offs (15, 20, 25 and 30 mm). The changes in sensitivity and specificity with changing cut-offs of cervical length are also graphically illustrated in Figure 4a.

Birth before 34 weeks' gestation (Table 2, Figure 5)

Four studies examined the performance of cervical length < 15 mm in predicting preterm birth at < 34 weeks' gestation (429 women, 78 cases of preterm birth (18.2%))[17–20]. Their pooled sensitivity was 46.2% (95% CI, 34.8–57.8%) and their pooled specificity was 93.7% (95% CI, 90.7–96.0%). The pooled LR+ and LR− estimates were 4.31 (95% CI, 2.73–6.82) and 0.63 (95% CI, 0.38–1.04), respectively.

The pooled sensitivity and specificity of cervical length < 20 mm for predicting preterm birth at < 34 weeks were 49.4% (95% CI, 37.9–60.9%) and 93.1% (95% CI, 89.7–95.7%), respectively (two studies, 385 women, 79 births at < 34 weeks (20.5%)), with high inconsistency across the two studies[19,22]. Cervical length < 25 mm (five studies, 735 women, 84 births at < 34 weeks (11.4%)) had a sensitivity of 64.3% (95% CI, 53.1–74.4%) and a specificity of 68.4% (95% CI, 64.6–71.9%)[18,19,27,35,36]. There was no heterogeneity in sensitivity ($I^2 = 0.0\%$), but it was very high in specificity ($I^2 = 95.2\%$). Figure 4b illustrates the changes in sensitivity and specificity for different cut-offs of cervical length in predicting birth before 34 weeks' gestation.

Sensitivity analysis

No sensitivity analysis was performed for study quality, as there were no obvious flaws in ultrasound methodology based on the descriptions given in the corresponding papers. For studies with recruitment exclusively at < 34 weeks and birth within 1 week from presentation

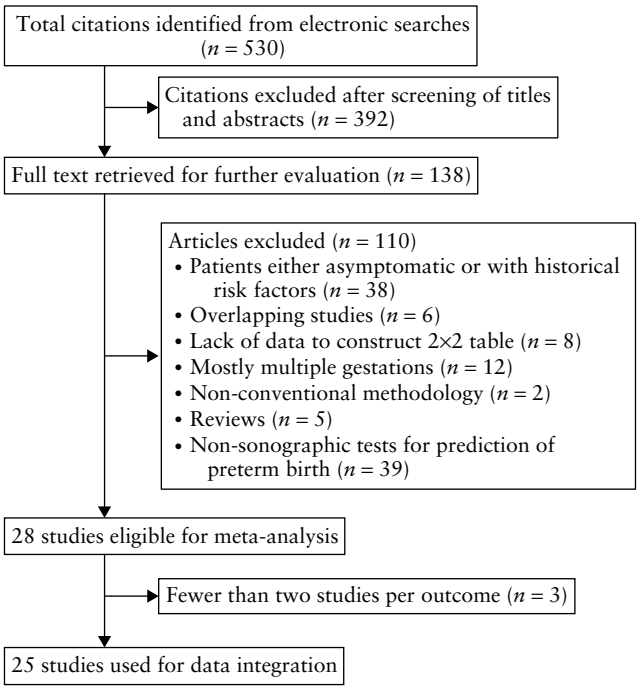

Figure 1 Flow diagram of search and selection processes.

Table 1 Characteristics of studies eligible for analysis

Study	*Description*	*Cut-off(s)/test(s)*	*Delivery*
Ultrasound assessment only			
Boudhraa *et al.* 2008[15]	56 symptomatic (uterine contractions proven by tocodynamometry, clinical or sonographic cervical changes) with singleton pregnancy at between 28 and 34 weeks. Women with acute fetal distress, retroplacental hematoma, intrauterine fetal death or severe fetal malformations were excluded.	CL < 25 mm	< 48 h, < 7 days
Schmitz *et al.* 2008[16]	395 symptomatic women (≥ 4 contractions/30 min, ≥ 30 s duration) with singleton pregnancy at between 24 and 34 + 6 weeks. Women with multiple pregnancy, PROM, cervical dilatation > 3 cm, cervical cerclage, uterine anomalies, placenta previa, placental abruption, IUGR, pre-eclampsia or medically indicated preterm delivery were excluded.	CL < 15, 20, 25, 30 mm	< 48 h, < 7 days, < 14 days, < 35 weeks, < 37 weeks
Alfirevic *et al.* 2007[17]	41 women with threatened PTL were randomized to have a transvaginal ultrasound ($n = 21$) or receive therapy as planned ($n = 20$). This is the subset allocated to TVS. Women with PROM, antenatal corticosteroids or oral steroid therapy were excluded.	CL < 15 mm	< 34 weeks
Gramellini *et al.* 2007[18]	108 symptomatic women (≥ 4 contractions/20 min) with singleton pregnancy at between 20 and 33 weeks. Women with PROM, dilatation ≥ 3 cm or cerclage were excluded.	CL < 15, 25 mm	< 7 days, < 34 weeks
Palacio *et al.* 2007[19]	333 symptomatic women with singleton pregnancy at between 24 and < 36 weeks (stratified in 2 groups: pre-32 weeks and post-32 weeks). Patients with Bishop score ≥ 6 or PROM and those who delivered within 24 h or iatrogenically were excluded.	CL < 15, 20, 25 mm	< 7 days, < 34 weeks
Holst *et al.* 2006[20]	87 symptomatic women (≥ 2 contractions/10 min for ≥ 30 min on CTG) with singleton pregnancy at between 22 + 0 and 33 + 6 weeks. Cases with PROM, PV bleeding, uterine or fetal malformations, imminent delivery, cervical cerclage or fetal distress were excluded. Amniocentesis was also performed to test for AF microbial invasion.	CL < 15 mm	< 7 days, < 34 weeks
Jenkins *et al.* 2006[21]	66 symptomatic women (no definition of symptoms) with singleton pregnancy at between 23 and 34 weeks and dilatation < 3 cm. Women with PROM or cervical cerclage were excluded. Dynamic cervical change was also assessed.	CL < 20 mm	< 37 weeks
Daskalakis *et al.* 2005[22]	172 symptomatic women with singleton pregnancies at between 24 and 34 weeks, with ≤ 2 cm dilatation and intact membranes.	CL < 20 mm	< 34 weeks
Gomez *et al.* 2005[23]	401 women with singleton pregnancies at between 25 and 35 weeks, with symptoms of preterm labor (contractions ≥ 3/30 min), dilatation ≤ 3 cm and intact membranes. Amniocentesis was also performed to test for AF microbial invasion.	CL < 15 mm	< 48 h, < 7 days, < 32 weeks, < 35 weeks
Kwasan *et al.* 2005[24]	69 symptomatic women with singleton pregnancy at between 24 and 35 + 6 weeks, with dilatation ≤ 3 cm and intact membranes. Cases with severe maternal or fetal complications (e.g. pre-eclampsia) or fetal distress necessitating immediate delivery were excluded.	CL < 30 mm	< 36 weeks
Tsoi *et al.* 2005[7]	510 symptomatic women (painful regular contractions) with singleton pregnancy at between 24 and 33 + 6 weeks. Women with dilatation > 3 cm and PROM were excluded.	CL < 5, 10, 15, 20 mm	< 48 h, < 7 days, < 35 weeks
Volumenie *et al.* 2004[25]	59 symptomatic women (contractions on tocodynamometry + cervical change) with singleton pregnancy, presenting at between 18 and 34 weeks, with dilatation ≤ 3 cm and intact membranes. Cases with cervical cerclage, pre-eclampsia, fetal pathology or maternal disease were excluded.	CL < 27 mm	< 37 weeks

Table 1 (*Continued*)

Study	*Description*	*Cut-off(s)/test(s)*	*Delivery*
Vendittelli *et al.* 2001[26]	200 women with singleton (87%) or twin (13%) pregnancy, hospitalized for preterm labor (contractions on external tocodynamometry at least 1/10 min and/or cervical changes), presenting at between 18 and 36 weeks. Women with cervical dilatation > 3 cm, PROM, cerclage, vaginal bleeding placenta previa, fetal malformation or delivery within 24 h from presentation were excluded.	CL < 30 mm	< 37 weeks
Crane *et al.* 1997[27]	136 symptomatic women with singleton pregnancy (+26 with twin) at between 23 and 33 weeks. Women with dilatation ≥ 3 cm or PROM were excluded. Cases with vaginal bleeding, placenta previa, ruptured membranes, cerclage, multiple gestation higher than twins or known stillbirth were excluded. The patients were examined after contractions had arrested.	CL < 25, 30 mm	< 34, 37 weeks
Gomez *et al.* 1994[28]	59 symptomatic women with singleton pregnancy at between 20 and 35 weeks, cervical dilatation < 3 cm and intact membranes.	CL < 18 mm	< 37 weeks
Murakawa *et al.* 1993[29]	32 symptomatic women (≥ 1 contraction/10 min) with singleton pregnancy at between 25 and 35 weeks. No data on membranes and dilatation. Women with PV bleeding may have been included.	CL < 20, 25 mm	< 37 weeks
Ultrasound plus biochemistry			
Eroglu *et al.* 2007[30]	51 symptomatic women (> 10 contractions/h) with singleton pregnancy at between 24 and 35 weeks. Women with dilatation ≥ 3 cm or PROM were excluded. Cases with vaginal bleeding, recent intercourse, uterine anomalies, fetal anomalies, abruption, IUGR or pre-eclampsia were also excluded.	CL < 20 mm; fFN pos/neg *ph*IGFBP-1 30 μg/L	< 7 days, < 35 weeks
Ness *et al.* 2007[31]	100 symptomatic women (≥ 6 contractions/h or symptoms suggestive) with singleton pregnancy at between 24 and 33 + 6 weeks, with intact membranes and dilatation < 3 cm. Cases with PV bleeding, recent cervical manipulation, uterine anomaly, triplets, cerclage and known short cervix were excluded.	CL < 20 mm; fFN pos/neg	< 37 weeks
Botsis *et al.* 2006[32]	62 singleton symptomatic (≥ 2 contractions/10 min) at between 24 and 26 weeks, with intact membranes and no signs of chorioamnionitis.	CL < 15 mm; plasma proMMP-9 67.2 ng/mL	< 36 weeks
Schmitz *et al.* 2006[33]	359 symptomatic women (≥ 4 contractions/30 min) with singleton pregnancy at between 18 and 34 + 6 weeks, with intact membranes and dilatation ≤ 3 cm. Cases with PV bleeding, recent cervical manipulation or intercourse, uterine anomalies, placenta previa, abruption, iatrogenic delivery, pre-eclampsia or IUGR were excluded.	CL < 15, 20, 25 mm fFN 50 ng/mL	

Table 1 (*Continued*)

Study	*Description*	*Cut-off(s)/test(s)*	*Delivery*
Tekesin *et al.* 2005[35]	117 symptomatic (4 contractions/20 min or ≥ 8/ h) women with singleton pregnancy presenting at between 24 and 34 weeks. Cases with PROM, cervical cerclage, cervical dilatation ≥ 3 cm, placenta previa, clinical criteria of intrauterine infection, vaginal bleeding of unknown origin, intrauterine growth restriction, pre-eclampsia, suspected fetal asphyxia, or a major fetal anomaly were excluded	CL < 25 mm; rapid fFN pos/neg; bacterial vaginosis; vaginal infection	< 34 weeks
Tekesin *et al.* 2005[34]	170 symptomatic women (4 contractions/20 min or 8/h) with singleton pregnancy at between 24 and 34 + 6 weeks, presenting with intact membranes and dilatation < 3 cm. Cases with cervical cerclage, placenta previa, signs of intrauterine infection, vaginal bleeding, IUGR, pre-eclampsia, suspected fetal asphyxia and fetal anomaly were excluded	CL < 25 mm; bacterial vaginosis; vaginal infection; fFN > 50 ng/mL	< 37 weeks
Nakai *et al.* 2005[36]	161 symptomatic women (≥ 6 contractions/h) with singleton pregnancy at between 22 and 28 weeks with intact membranes. Women with vaginal bleeding, uterine anomaly, chorioamnionitis, fetal anomalies or placental abruption were excluded. Women with cervical cerclage were included.	CL < 25 mm; elastase 1.6 mg/L; bacterial vaginosis	< 34 weeks
Gonzalez *et al.* 2004[37]	192 symptomatic (regular uterine contractions with cervical change) women with singleton pregnancy presenting before 34 + 0 weeks. Women with cerclage or maternal/fetal pathology indicating induction of delivery were excluded.	CL < 30 mm; fFN 50 ng/mL	< 34, 37 weeks
Hincz *et al.* 2002[38]	82 symptomatic women (≥ 4 contractions/h with cervical changes) with singleton pregnancy at between 24 and 34 weeks, intact membranes and dilatation ≤ 3 cm. Cases with PV bleeding, intrauterine infection, placenta previa, IUGR or pre-eclampsia were excluded.	CL < 20 mm; fFN 50 ng/mL	< 28 days
Kurkinen-Raty *et al.* 2001[39]	77 symptomatic (not defined) with singleton pregnancy or asymptomatic women with shortening cervix at between 22 and 32 weeks, with intact membranes. Women with ongoing preterm labor were excluded.	CL < 29.3 mm; CV IL-6 61 ng/L; CV IL-8 3739 ng/L; CV phIGFBP-1 21 mg/L	< 37 weeks
Rozenberg *et al.* 1997[40]	76 symptomatic women with signs of premature labor requiring hospitalization (≥ 2 contractions/10 min, shortening or dilatation of the cervix) at between 24 and 34 weeks and dilatation ≤ 2 cm. Cases with multiple pregnancy, PROM, cervical cerclage, placenta previa, placental abruption and medically indicated preterm delivery were excluded.	< 26 mm; fFN 50 ng/mL	< 37 weeks
Rizzo *et al.* 1996[41]	108 symptomatic women (≥ 4 contractions/20 min or ≥ 8/h) with singleton pregnancy at between 24 and 36 weeks with intact membranes and dilatation ≤ 3 cm. Cases with maternal or fetal complications were excluded.	CL < 20 mm; fFN > 60 ng/mL	< 37 weeks

AF, amniotic fluid; CL, cervical length; CTG, cardiotocography: CV, cervicovaginal; fFN, fetal fibronectin; Glc, glucose; IL, interleukin; IUGR, intrauterine growth restriction; *ph*IGFBP-1, phosphorylated insulin-like growth factor binding protein-1; PROM, preterm rupture of membranes; proMMP-9: pro-matrix metalloproteinase-9; PTL, preterm labor; PV, per vagina; TVS, transvaginal sonography; WBC, white blood cells.

Table 2 Performance estimates for prediction of preterm birth in symptomatic women

Cut-off for test	*48 h*	*7 days*	*32 weeks*	*34 weeks*	*35 weeks*	*36 weeks*	*37 weeks*
Cervical length < 15 mm	Gomez 2005[23] Tsoi 2005[7] Schmitz 2008[16]	Gomez 2005[23] Holst 2006[20] Tsoi 2005[7] Palacio 2007[19] Gramellini 2007[18] Schmitz 2006[33]	Gomez 2005[23]	Holst 2006[20] Palacio 2007[19] Alfirevic 2007[17] Gramellini 2007[18]	Gomez 2005[23] Tsoi 2005[7] Schmitz 2006[33]	Botsis 2006[32]	
Sensitivity (%)	**71.1 (59.5–80.9)**	**59.9 (52.7–66.8)**	**63.0 (44.2–78.5)**	**46.2 (34.8–57.8)**	**55.3 (48.3–62.3)**	81.8 (52.3–94.9)	
Specificity (%)	**86.6 (84.6–88.5)**	**90.5 (89.0–91.9)**	**91.7 (86.4–95.1)**	**93.7 (90.7–96.0)**	**90.9 (89.0–92.6)**	92.2 (81.5–97.5)	
Number of women	**1266**	**1781**		**429**	**1253**		
I^2 sens/I^2 spec	**88.7%/57.2%**	**91.0%/87.2%**		**88.9%/93.7%**	**83.8%/59.1%**		
LR+		**5.71 (3.77–8.65)**		**4.31 (2.73–6.82)**	**6.14 (4.15–9.09)**		
LR−		**0.51 (0.33–0.80)**		**0.63 (0.38–1.04)**	**0.49 (0.34–0.71)**		
Cervical length < 20 mm	Tsoi 2005[7] Schmitz 2008[16]	Tsoi 2005[7] Palacio 2007[19] Eroglu 2007[30] Schmitz 2006[33]		Palacio 2007[19] Daskalakis 2005[22]	Tsoi 2005[7] Eroglu 2007[30] Schmitz 2006[33]		Ness 2007[31] Jenkins 2006[21] Tekesin 2005[34] Rizzo 1996[41] Murakawa 1993[29] Gomez 1994**[28]
Sensitivity (%)	**86.8 (71.9–95.6)**	**75.4 (66.6–82.9)**		**49.4 (37.9–60.9)**	**69.0 (60.3–76.8)**		**57.3 (48.8–65.6)**
Specificity (%)	**72.2 (69.1–75.2)**	**79.6 (77.1–81.9)**		**93.1 (89.7–95.7)**	**77.1 (74.0–80.0)**		**80.9 (75.9–85.2)**
Number of women	**905**	**1263**		**385**	**915**		**441**
I^2 sens/I^2 spec	**88.9%/78.6%**	**88.4%/94.7%**		**85.2%/65.0%**	**45.6%/91.5%**		**74.3%/78.2%**
LR+		**3.74 (2.77–5.05)**		**6.04 (0.85–43.1)**	**2.98 (2.13–4.18)**		**2.70 (2.07–3.52)**
LR−		**0.33 (0.15–0.73)**		**0.63 (0.31–1.28)**	**0.41 (0.31–0.56)**		**0.57 (0.43–0.77)**
Cervical length < 25 mm	Boudhraa 2008[15] Schmitz 2008[16]	Boudhraa 2008[15] Palacio 2007[19] Gramellini 2007[18] Schmitz 2006[33]		Palacio 2007[19] Gramellinzi 2007[18] Nakai 2005[36] Tekesin 2005[35] Crane 1997[27]	Schmitz 2006[33]		Tekesin 2005[34] Volumenie 2004[25] Crane 1997[27] Rozenberg 1997[40] Murakawa 1993[29]
Sensitivity (%)	**88.0 (68.8–97.5)**	**78.3 (67.9–86.6)**		**64.3 (53.1–74.4)**	**75.0 (61.2–85.1)**		**71.2 (62.7–78.8)**
Specificity (%)	**58.9 (54.1–63.6)**	**70.8 (67.4–74.0)**		**68.4 (64.6–71.9)**	**63.0 (57.5–68.2)**		**69.8 (64.2–75.0)**
Number of women	**451**	**856**		**735**			**430**
I^2 sens/I^2 spec	**60.0%/0.0%**	**34.4%/89.8%**		**0.0%/95.2%**			**0.0%/71.4%**
LR+	**2.09 (1.70–2.57)**	**2.77 (2.15–3.59)**		**2.22 (1.43–3.44)**			**2.24 (1.71–2.93)**
LR−	**0.26 (0.10–0.68)**	**0.33 (0.22–0.50)**		**0.54 (0.41–0.72)**			**0.43 (0.33–0.57)**
Cervical length < 30 mm	Schmitz 2008[16]	Schmitz 2008[16]		Gonzalez 2004[37] Crane 1997[27]		Kwasan 2005[24]	Gonzalez 2004[37] Kurkinen-Raty 2001[39] Vendittelli 2001[26] Crane 1997[27]
Sensitivity (%)	88.2 (65.7–96.7)	93.8 (79.9–98.3)		**93.8 (82.8–98.7)**		93.3 (78.7–98.1)	**83.4 (77.7–88.2)**
Specificity (%)	40.0 (35.1–45.0)	41.9 (36.9–47.0)		**42.1 (36.3–48.2)**		82.1 (67.3–91.0)	**52.5 (47.6–57.3)**
Number of women				**328**			**634**
I^2 sens/I^2 spec				**59.9%/94%**			**0.0%/88.7%**
LR+				**1.57 (1.19–2.07)**			**1.77 (1.34–2.34)**
LR−				**0.20 (0.06–0.66)**			**0.35 (0.25–0.48)**

Only the first author's name of each study is given. Numbers in parentheses are 95% CIs. Numbers in bold are pooled estimates. LR+, positive likelihood ratio; LR−, negative likelihood ratio.

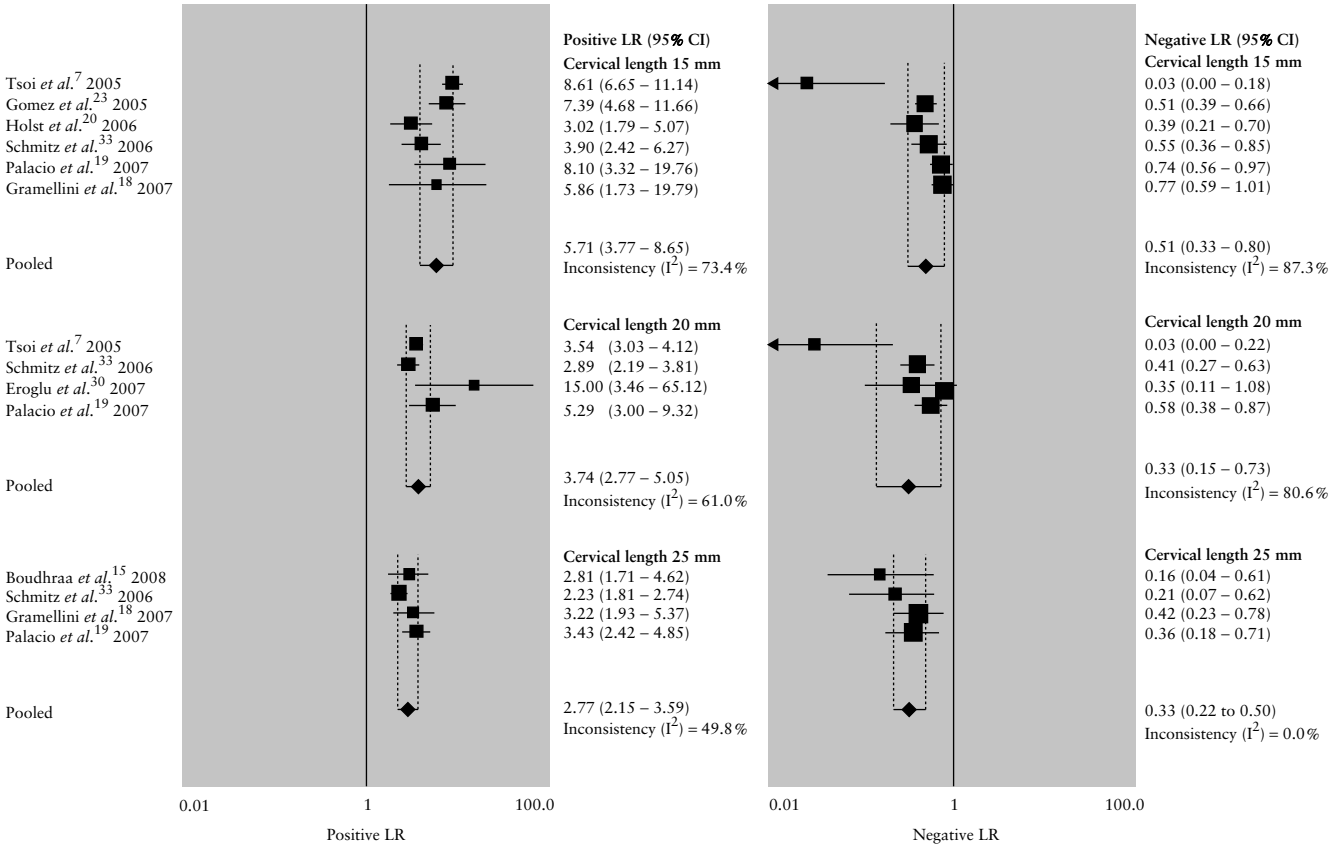

Figure 2 Positive and negative likelihood ratios for birth within 1 week from presentation.

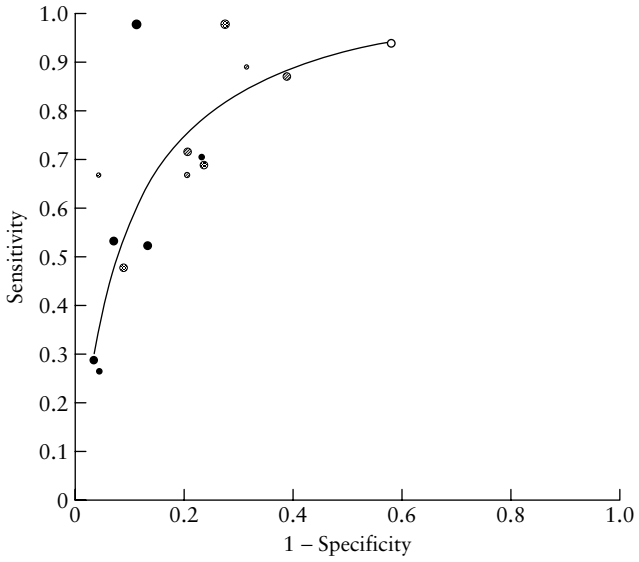

Figure 3 Composite summary receiver–operating characteristics (SROC) curve for birth within 7 days from presentation, including all tested cut-offs for cervical length. Each circle represents data from a specific study on a specific cut-off. ●, Cervical length (CL) < 15 mm; ⊘, CL < 20 mm; ⊗, CL < 25 mm; ○, CL < 30 mm. Size of dots varies according to the 'weight' each study has for the calculation of the curve. The values for AUC, Q* and SE(Q*) are 0.429, 0.775 and 0.018, respectively. These values were calculated for the range of available data. Extrapolation of the SROC curve would yield an AUC of 0.841.

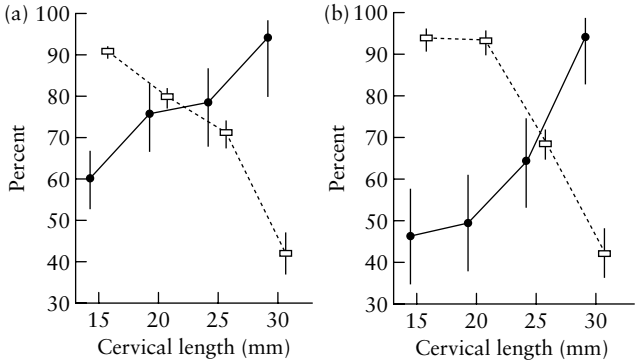

Figure 4 Changes in sensitivity (●) and specificity (▭) for different cut-offs of cervical length in predicting birth within 1 week from presentation (a) and before 34 weeks' gestation (b).

(four studies, 821 women, 93 women with birth within 1 week (11.3%)), the sensitivity was 71.0% (95% CI, 60.6–79.9%) and the specificity was 89.8% (95% CI, 87.4–91.9%)[7,18–20]. Heterogeneity remained profound ($I^2 > 85\%$). The pooled LR+ and LR− estimates were 5.19 (95% CI, 2.29–11.74) and 0.38 (95% CI, 0.11–1.34), respectively.

DISCUSSION

Severe prematurity is the leading cause of perinatal morbidity and mortality, and much effort is put into

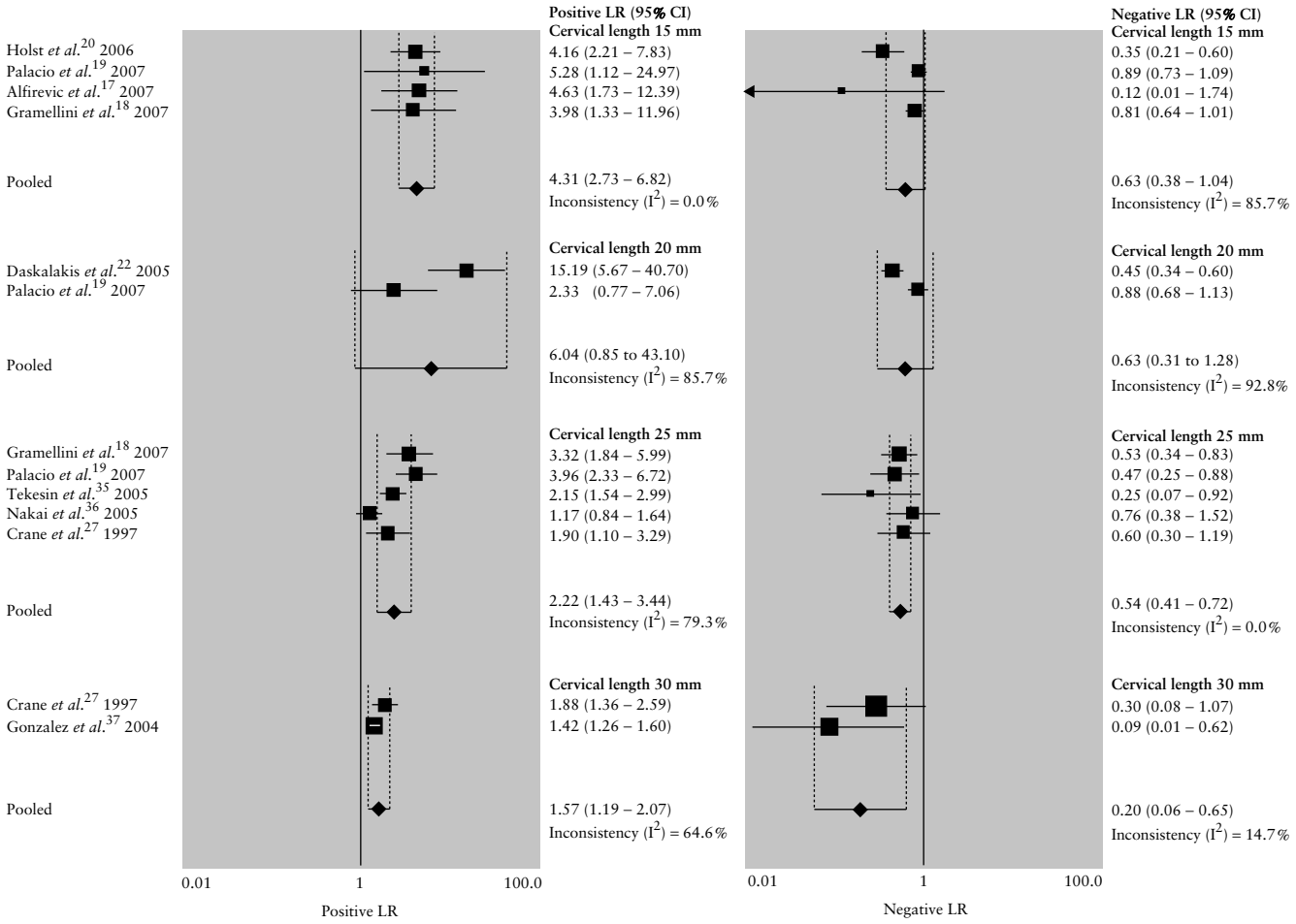

Figure 5 Positive and negative likelihood ratios for birth before 34 weeks' gestation.

detecting and treating mothers who are likely to deliver very early. In contrast to asymptomatic women, for whom screening measurement of the cervical length in the midtrimester can accurately detect many who will subsequently deliver preterm[8], and for whom prophylactic administration of progesterone may effectively prevent extreme prematurity in many of them[9], both detection and prevention of true preterm birth appear to be problematic in women who present with symptoms of preterm labor.

The elusive diagnosis of preterm labor

Preterm labor is largely a subjective diagnosis, and the prevalence of actual preterm birth in women with symptoms of preterm labor shows great variation between studies. A flexible definition of preterm labor most probably underlies the fact that, of women presenting with threatened preterm labor, only 7% actually delivered within 48 h from presentation (data from 1266 women)[7,16,23] and only 11% actually delivered within 1 week (data from 1781 women)[7,18–20,23,33]. When confining our analysis to studies with both recruitment at or before 34 weeks' gestation and explicit criteria for preterm labor (certain frequency and duration of contractions ± tocographic ascertainment)[7,18,20], the prevalence of birth within 1 week slightly increases as expected, but again it is only 12.6%, and still varied greatly across the studies, ranging from 8.4%[7] to 31.0%[20].

Predictive performance of cervical length

We used random effects models, which, in the presence of heterogeneity, are more conservative, providing wider confidence intervals; nevertheless, our results must be taken cautiously. Given these limitations, a cervical length of < 15 mm is seen in less than 10% of symptomatic women and predicts approximately 60% (or 70% if the patient presents before 34 weeks' gestation) of the women who will deliver within the following week. A short cervix (< 15 mm) increases the chances of delivering within 1 week by 5.7 times. Conversely, a longer cervix can be a reassuring sign, with the vast majority of such women not delivering within the following week. The pooled negative predictive value (NPV, i.e. the likelihood not to deliver if the test is normal) was 96% for presentation before 34 weeks and CL > 15 mm, i.e. only 4% of symptomatic women presenting before 34 weeks would deliver within the following week if the cervix was longer than 15 mm. Our analysis showed that a cervical length of 15 mm is sufficient to exclude birth within a week

in 95–96% of cases, and a higher cut-off would not essentially increase the NPV. Based on our weighted estimates, for the pooled prevalence of 11.1% for birth within 1 week from presentation, the NPVs of 15 mm, 20 mm and 25 mm would be 94.8%, 96.3% and 95.8%, respectively.

One should always be cautious when interpreting NPVs, as they are so strongly dependent on the prevalence of the condition of interest that even a test with no diagnostic value (i.e. sensitivity 50% and specificity 50%) would yield an NPV of 95% if the prevalence of the condition is 5%. Therefore, NPVs should be interpreted only in their broad sense or they should be used for providing a means of comparison between different predictive methods for a given prevalence of the condition of interest. Given these limitations, testing with cervical length measurement still appears to assist rationalizing clinical management, provided that the clinician interprets the results appropriately.

Strengths and limitations of our analysis: heterogeneity

The performance of cervical length measurement for the prediction of preterm birth in symptomatic women has been systematically reviewed before, either as a stand-alone review of studies in symptomatic women[42], or as a subset of a wider meta-analysis including both symptomatic and asymptomatic women[43,44]. In comparison with these publications, our analysis has included much larger samples, as many of the largest studies have been published within the last 5 years.

Apart from the results on the predictive performance of cervical ultrasound, an important methodological issue highlighted in our analysis was the presence of considerable heterogeneity across studies in most pooled estimates of sensitivity and specificity. In cases with such extreme inconsistency, the pooled estimates, despite their narrow confidence intervals, are often misleading unless the potential causes for the underlying clinical heterogeneity are properly explored[45]. An obvious cause of clinical heterogeneity would lie in the inclusion criteria for women in the individual studies, i.e. in the definition of preterm labor. Variation in defining what is preterm labor in a patient who presents with suggestive complaints is reflected in the variation of true preterm birth rate across studies as discussed above; however, it still does not explain the inconsistency in sensitivity and specificity, which would be independent of variations in the prevalence of preterm birth. In studies with cervical assessment, different methodology and skill levels may affect its prognostic performance, and data on scan methodology were therefore recorded during quality assessment in our analysis. All studies appeared to follow the same good methodology, and the actual performance of individual operators could not be quantified from information available in the published articles. Also, there was no issue of different clinical cut-offs, as our analytical groups were based on this particular criterion. Therefore, there were no obvious sources of clinical heterogeneity that would explain the considerable inconsistency in the estimates. A similar finding, although with a smaller sample size, has also been reported in the meta-analysis of Honest *et al.*[43], where quality could not provide an explanation for the variation in accuracy among subgroups of studies based on gestational age at testing, test thresholds and reference standards. This raises the question of methodological flaws that either were not reported at all or were under-reported in some studies.

REFERENCES

1. Moutquin JM. Classification and heterogeneity of preterm birth. *BJOG* 2003; **110** (Suppl 20): 30–33.
2. Hack M, Fanaroff AA. Outcomes of children of extremely low birthweight and gestational age in the 1990s. *Semin Neonatol* 2000; **5**: 89–106.
3. Hosono S, Ohno T, Kimoto H, Shimizu M, Harada K. Morbidity and mortality of infants born at the threshold of viability: ten years' experience in a single neonatal intensive care unit, 1991–2000. *Pediatr Int* 2006; **48**: 33–39.
4. Marret S, Ancel PY, Marpeau L, Marchand L, Pierrat V, Larroque B, Foix-L'Hélias L, Thiriez G, Fresson J, Alberge C, Rozé JC, Matis J, Bréart G, Kaminski M; Epipage Study Group. Neonatal and 5-year outcomes after birth at 30–34 weeks of gestation. *Obstet Gynecol* 2007; **110**: 72–80.
5. Scott CL, Chavez GF, Atrash HK, Taylor DJ, Shah RS, Rowley D. Hospitalizations for severe complications of pregnancy, 1987–1992. *Obstet Gynecol* 1997; **90**: 225–229.
6. Bacak SJ, Callaghan WM, Dietz PM, Crouse C. Pregnancy-associated hospitalizations in the United States, 1999–2000. *Am J Obstet Gynecol* 2005; **192**: 592–597.
7. Tsoi E, Fuchs IB, Rane S, Geerts L, Nicolaides KH. Sonographic measurement of cervical length in threatened preterm labor in singleton pregnancies with intact membranes. *Ultrasound Obstet Gynecol* 2005; **25**: 353–356.
8. To MS, Skentou CA, Royston P, Yu CK, Nicolaides KH. Prediction of patient-specific risk of early preterm delivery using maternal history and sonographic measurement of cervical length: a population-based prospective study. *Ultrasound Obstet Gynecol* 2006; **27**: 362–367.
9. Fonseca EB, Celik E, Parra M, Singh M, Nicolaides KH. Progesterone and the risk of preterm birth among women with a short cervix. *N Engl J Med* 2007; **357**: 462–469.
10. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997; **127**: 820–826.
11. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994; **271**: 703–707.
12. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**: 557–560.
13. Heath VC, Southall TR, Souka AP, Novakov A, Nicolaides KH. Cervical length at 23 weeks of gestation: relation to demographic characteristics and previous obstetric history. *Ultrasound Obstet Gynecol* 1998; **12**: 304–311.
14. 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**: 859–867.
15. Boudhraa K, Rahouej H, Gara MF. [Transvaginal ultrasound of the cervix in the estimation of severity of premature labour.] [Article in French] *Tunis Med* 2008; **86**: 745–748.

16. Schmitz T, Kayem G, Maillard F, Lebret MT, Cabrol D, Goffinet F. Selective use of sonographic cervical length measurement for predicting imminent preterm delivery in women with preterm labor and intact membranes. *Ultrasound Obstet Gynecol* 2008; **31**: 421–426.
17. Alfirevic Z, Allen-Coward H, Molina F, Vinuesa CP, Nicolaides K. Targeted therapy for threatened preterm labor based on sonographic measurement of the cervical length: a randomized controlled trial. *Ultrasound Obstet Gynecol* 2007; **29**: 47–50.
18. Gramellini D, Fieni S, Kaihura C, Modena AB. Cervical length as a predictor of preterm delivery: gestational age-related percentiles vs fixed cut-offs. *Acta Biomed* 2007; **78**: 220–224.
19. Palacio M, Sanin-Blair J, Sanchez M, Crispi F, Gómez O, Carreras E, Coll O, Cararach V, Gratacós E. The use of a variable cut-off value of cervical length in women admitted for preterm labor before and after 32 weeks. *Ultrasound Obstet Gynecol* 2007; **29**: 421–426.
20. Holst RM, Jacobsson B, Hagberg H, Wennerholm UB. Cervical length in women in preterm labor with intact membranes: relationship to intra-amniotic inflammation/microbial invasion, cervical inflammation and preterm delivery. *Ultrasound Obstet Gynecol* 2006; **28**: 768–774.
21. Jenkins SM, Kurtzman JT, Osann K. Dynamic cervical change: is real-time sonographic cervical shortening predictive of preterm delivery in patients with symptoms of preterm labor? *Ultrasound Obstet Gynecol* 2006; **27**: 373–376.
22. Daskalakis G, Thomakos N, Hatziioannou L, Mesogitis S, Papantoniou N, Antsaklis A. Cervical assessment in women with threatened preterm labor. *J Matern Fetal Neonatal Med* 2005; **17**: 309–312.
23. Gomez R, Romero R, Nien JK, Chaiworapongsa T, Medina L, Kim YM, Yoon BH, Carstens M, Espinoza J, Iams JD, Gonzalez R. A short cervix in women with preterm labor and intact membranes: a risk factor for microbial invasion of the amniotic cavity. *Am J Obstet Gynecol* 2005; **192**: 678–689.
24. Kwasan S, Paisarntuntiwong R, Charoenchainont P. Cervical length measurement by transvaginal sonography in preterm pregnant women for prediction of preterm birth. *J Med Assoc Thai* 2005; **88** (Suppl 2): S48–S55.
25. 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**: 33–37.
26. Vendittelli F, Mamelle N, Munoz F, Janky E. Transvaginal ultrasonography of the uterine cervix in hospitalized women with preterm labor. *Int J Gynaecol Obstet* 2001; **72**: 117–125.
27. Crane JM, Van den Hof M, Armson BA, Liston R. Transvaginal ultrasound in the prediction of preterm delivery: singleton and twin gestations. *Obstet Gynecol* 1997; **90**: 357–363.
28. Gomez R, Galasso M, Romero R, Mazor M, Sorokin Y, Gonçalves L, Treadwell M. 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**: 956–964.
29. Murakawa H, Utumi T, Hasegawa I, Tanaka K, Fuzimori R. Evaluation of threatened preterm delivery by transvaginal ultrasonographic measurement of cervical length. *Obstet Gynecol* 1993; **82**: 829–832.
30. Eroglu D, Yanik F, Oktem M, Zeyneloglu HB, Kuscu E. Prediction of preterm delivery among women with threatened preterm labor. *Gynecol Obstet Invest* 2007; **64**: 109–116.
31. Ness A, Visintine J, Ricci E, Berghella V. Does knowledge of cervical length and fetal fibronectin affect management of women with threatened preterm labor? A randomized trial. *Am J Obstet Gynecol* 2007; **197**: 426 e421–427.
32. Botsis D, Makrakis E, Papagianni V, Kouskouni E, Grigoriou O, Dendrinos S, Creatsas G. The value of cervical length and plasma proMMP-9 levels for the prediction of preterm delivery in pregnant women presenting with threatened preterm labor. *Eur J Obstet Gynecol Reprod Biol* 2006; **128**: 108–112.
33. Schmitz T, Maillard F, Bessard-Bacquaert S, Kayem G, Fulla Y, Cabrol D, Goffinet F. Selective use of fetal fibronectin detection after cervical length measurement to predict spontaneous preterm delivery in women with preterm labor. *Am J Obstet Gynecol* 2006; **194**: 138–143.
34. Tekesin I, Eberhart LH, Schaefer V, Wallwiener D, Schmidt S. Evaluation and validation of a new risk score (CLEOPATRA score) to predict the probability of premature delivery for patients with threatened preterm labor. *Ultrasound Obstet Gynecol* 2005; **26**: 699–706.
35. Tekesin I, Wallwiener D, Schmidt S. The value of quantitative ultrasound tissue characterization of the cervix and rapid fetal fibronectin in predicting preterm delivery. *J Perinat Med* 2005; **33**: 383–391.
36. Nakai A, Taniuchi Y, Miyake H, Nakai M, Yokota A, Takeshita T. Increased level of granulocyte elastase in cervical secretion is an independent predictive factor for preterm delivery. *Gynecol Obstet Invest* 2005; **60**: 87–91.
37. Gonzalez N, Bige V, Kandoussi S, Graesslin O, Quereux C, Gabriel R. [Ultrasonographic measurement of cervical length in twin pregnancies with preterm labor: comparison with singleton pregnancies.] [Article in French] *Gynecol Obstet Fertil* 2004; **32**: 122–127.
38. Hincz P, Wilczynski J, Kozarzewski M, Szaflik K. Two-step test: the combined use of fetal fibronectin and sonographic examination of the uterine cervix for prediction of preterm delivery in symptomatic patients. *Acta Obstet Gynecol Scand* 2002; **81**: 58–63.
39. Kurkinen-Raty M, Ruokonen A, Vuopala S, Koskela M, Rutanen EM, Kärkkäinen T, Jouppila P. Combination of cervical interleukin-6 and -8, phosphorylated insulin-like growth factor-binding protein-1 and transvaginal cervical ultrasonography in assessment of the risk of preterm birth. *BJOG* 2001; **108**: 875–881.
40. Rozenberg P, Goffinet F, Malagrida L, Giudicelli Y, Perdu M, Houssin I, Safe-Femme, Nisand I. Evaluating the risk of preterm delivery: a comparison of fetal fibronectin and transvaginal ultrasonographic measurement of cervical length. *Am J Obstet Gynecol* 1997; **176**: 196–199.
41. Rizzo G, Capponi A, Arduini D, Lorido C, Romanini C. The value of fetal fibronectin in cervical and vaginal secretions and of ultrasonographic examination of the uterine cervix in predicting premature delivery for patients with preterm labor and intact membranes. *Am J Obstet Gynecol* 1996; **175**: 1146–1151.
42. Vendittelli F, Volumenie J. Transvaginal ultrasonography examination of the uterine cervix in hospitalised women undergoing preterm labour. *Eur J Obstet Gynecol Reprod Biol* 2000; **90**: 3–11.
43. Honest H, Bachmann LM, Coomarasamy A, Gupta JK, Kleijnen J, Khan KS. Accuracy of cervical transvaginal sonography in predicting preterm birth: a systematic review. *Ultrasound Obstet Gynecol* 2003; **22**: 305–322.
44. Leitich H, Brunbauer M, Kaider A, Egarter C, Husslein P. Cervical length and dilatation of the internal cervical os detected by vaginal ultrasonography as markers for preterm delivery: A systematic review. *Am J Obstet Gynecol* 1999; **181**: 1465–1472.
45. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 1994; **309**: 1351–1355.

