TRANSVAGINAL SONOGRAPHIC CERVICAL LENGTH FOR THE PREDICTION OF SPONTANEOUS PRETERM BIRTH IN TWIN PREGNANCIES: A SYSTEMATIC REVIEW AND METAANALYSIS

Agustín CONDE-AGUDELO, MD, MPH[a], **Roberto ROMERO, MD**[a,b,c], **Sonia S. HASSAN, MD**[a,b], and **Lami YEO, MD**[a,b]

[a]Perinatology Research Branch, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development/National Institutes of Health/Department of Health and Human Services, Bethesda, MD and Detroit, MI

[b]Department of Obstetrics and Gynecology, Wayne State University, Detroit, Michigan, USA

[c]Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI

Abstract

Objective—To assess the accuracy of transvaginal sonographic cervical length (CL) in predicting spontaneous preterm birth in women with twin pregnancies.

Study design—Systematic review and metaanalysis of predictive test accuracy.

Results—Twenty-one studies (16 in asymptomatic women and 5 in symptomatic women) with a total of 3523 women met the inclusion criteria. Among asymptomatic women, a CL ≤20 mm 20-24 weeks' gestation was the most accurate in predicting preterm birth <32 and <34 weeks' gestation (pooled sensitivities, specificities, and positive and negative likelihood ratios of 39% and 29%, 96% and 97%, 10.1 and 9.0, and 0.64 and 0.74, respectively). A CL ≤25 mm 20-24 weeks' gestation had a pooled positive likelihood ratio of 9.6 to predict preterm birth <28 weeks' gestation. The predictive accuracy of CL for preterm birth was low in symptomatic women.

Conclusion—Transvaginal sonographic CL 20-24 weeks' gestation is a good predictor of spontaneous preterm birth in asymptomatic women with twin pregnancies.

Keywords

Cervical length; preterm birth; prediction; metaanalysis; systematic review

INTRODUCTION

Despite advancing knowledge of the risk factors and mechanisms associated with preterm labor and delivery, the preterm birth rate has risen 36% in the United States during the last quarter century (from 9.4% in 1981 to 12.8% in 2006).[1] This increase has been explained in part by a rise in the number of indicated preterm births in singleton gestations and preterm delivery of multiple pregnancies that occurred as a result of assisted reproductive technologies.[2] Furthermore, the twin birth rate has risen 70% from 1980 (18.9 per 1000 live births) to 2006 (32.1 per 1000 live births).[1]

Corresponding author Dr Roberto Romero, Perinatology Research Branch, Intramural Division, NICHD/NIH/DHHS, Hutzel Women's Hospital, Box # 4, 3990 John R, Detroit, MI 48201, Telephone: +1 313 993 2700, Fax: +1 313 993 2694, prbchiefstaff@med.wayne.edu.

In the United States, the rates of preterm birth <37 and <32 weeks of gestation for twin pregnancies (60.4% and 12.1%, respectively) were 5.4 and 7.6 times the rates for singleton pregnancies (11.1% and 1.6%, respectively).[1] Overall, twin pregnancies comprise 15% of all preterm births[1] accounting for a disproportionate share of preterm births. Therefore, there is an urgent need to develop cost-effective tests for the prediction of preterm birth in twin pregnancies. The ability to identify women at high risk for spontaneous preterm birth could allow for patients to undergo targeted interventions such as transfer to a tertiary care center, antenatal corticosteroid administration and tocolysis, which might improve perinatal outcomes among twins. Previous reviews have suggested that transvaginal sonographic assessment of cervical length (CL) is an effective tool for predicting preterm birth, particularly in asymptomatic women or those at a higher risk of spontaneous preterm birth.[3-5] However, these reviews largely evaluated the use of CL in singleton pregnancies. In addition, published studies on predictive accuracy of CL for preterm birth in twin pregnancies report conflicting conclusions on the value of this test.

The objective of this study was to assess the value of transvaginal sonographic CL for the prediction of spontaneous preterm birth in women with twin pregnancies through the use of formal methods for systematic reviews and metaanalytic techniques.

MATERIALS AND METHODS

This systematic review was conducted following a prospectively prepared protocol and reported using recently recommended guidelines for systematic reviews of diagnostic test accuracy.[6]

Literature search

Electronic searches, without language restrictions, were performed in the MEDLINE (January 1966-November 2009), EMBASE (January 1980- November 2009), CINAHL (January 1982- November 2009), LILACS (January 1982-November 2009), and Medion (January 1974-November 2009) databases to identify potentially eligible studies. We applied the following algorithm both in Medical Subject Headings and in free-text words in MEDLINE: {cervical length OR [(transvaginal OR vaginal OR cervix OR cervical) AND (ultrasound OR ultrasonography OR ultrasonographic OR sonography OR sonographic)]} AND (preterm OR premature). This search strategy was also used for the other databases, adjusted according to specific requirements for the particular electronic database. Proceedings of the Society for Maternal-Fetal Medicine and international meetings on preterm birth and twin or multiple pregnancy, reference lists of identified studies, textbooks, and previously published systematic reviews were also searched. In addition, we contacted experts in the field to obtain unpublished studies. All searches were conducted independently by two of the authors (A.C-A and R.R) and results were merged.

Inclusion criteria

Studies were included if they met the following minimal criteria: (1) a cohort or cross-sectional study that evaluated the accuracy of transvaginal sonographic CL measurement to predict spontaneous preterm birth in asymptomatic or symptomatic pregnant women with twin pregnancies; (2) the outcome measures included any category of spontaneous preterm birth <37 weeks of gestation; (3) the studies provided the necessary information to generate 2×2 tables; and (4) the women had no therapeutic intervention resulting from the test result. When a study based its results on mixed (singleton and twin) pregnancies, unless data for twins were extractable separately, it was not considered for inclusion in the review. In cases of duplicate publication we selected the most recent and complete versions and supplemented if additional information appeared in the other publications

Studies were excluded from the systematic review if they: (1) were case-control studies because these tend to overestimate the predictive or diagnostic accuracy of a test;[7] (2) did not provide data on predictive estimates and sufficient information to calculate them could not be retrieved; or (3) they included women with cervical cerclage, previous cervical surgery, or premature rupture of membranes.

Studies deemed suitable were retrieved and reviewed independently by two authors (A.C-A and R.R) to determine inclusion. Disagreements were resolved by consensus.

Quality assessment

Methodological quality of included studies was assessed independently by two reviewers (A.C-A and R.R) using 4 of the 14 items of the Quality Assessment of Diagnostic Accuracy Studies(QUADAS) tool.[8] The remaining 10 QUADAS items were not used because they were not relevant to our review. Each item was scored as "yes", "no", or "unclear."

The items of the QUADAS tool evaluated and their interpretation were as follows:

1. Representative spectrum of patients. This item was scored "yes" when pregnant women with twin pregnancies were consecutively selected in a prospective way. Convenience sampling, such as arbitrary recruitment or nonconsecutive recruitment, was scored as "no."

2. Description of the test: This item was scored as "yes" if the study described sufficient details of the technique used for measuring CL such as plane in which images were obtained, anatomic references for the determination of CL, and number of measurements. If this information was not reported, then this item was scored as "no."

3. Blinding of index test result: This item was scored "yes" if the study clearly stated that clinicians treating the patient did not have knowledge of the CL results. If this did not appear to be the case, this item was scored as "no."

4. Reporting of study withdrawals. If there were withdrawals from the study, this item was scored as "yes" if withdrawals were explained or if a flow diagram of study participants was reported. If it appeared that some of the participants did not complete the study and these patients were not accounted for, then this item was scored as "no."

If there was insufficient information available to make a judgment of these items, then they were scored as "unclear." Summary score estimating the overall quality of an article was not calculated because the interpretation of such summary scores is problematic and potentially misleading.[9]

Discrepancies in quality assessment were resolved by discussion.

Data extraction

Data extraction was performed independently by 2 investigators (A.C-A and R.R) and recorded on a standardized form. There was no blinding of authorship. Information was extracted on study characteristics, study quality, and participant characteristics. With regard to transvaginal ultrasonography, we extracted data on the technique used for measuring CL, gestational age(s) at testing, and cutoff values used. Gestational age at testing was divided into 3 groups: <20, 20-24, and >24 weeks of gestation. For studies in which the reported gestational age at testing encompassed ≥2 of the groups, we classified them according to their mean gestational age at testing. For each study, for all reported cutoff values for CL, and for all categories of spontaneous preterm birth, we then extracted the number of true-

positive, false-positive, true-negative, and false-negative results. When predictive accuracy data were not available, we recalculated them from the reported results.

Studies reporting on spontaneous preterm birth <35 weeks of gestation were considered with those reporting spontaneous preterm birth <34 weeks of gestation, as both these gestational ages have similar neonatal outcomes. In the same way, studies reporting spontaneous preterm birth <36 weeks of gestation were considered with those reporting spontaneous preterm birth <37 weeks of gestation. We extracted data separately for asymptomatic women and for women with threatened preterm labor.

Disagreements in data extraction were resolved by discussion among authors.

Statistical analysis

For both asymptomatic and symptomatic women, we synthesized data for spontaneous preterm birth at <34 and <37 weeks of gestation. In addition, for asymptomatic women we synthesized data for spontaneous preterm birth at <28 and <32 weeks of gestation.

Data extracted from each study were arranged in 2×2 contingency tables. When these tables contained cells for which the value was 0, we added 0.5 to those cells to allow for the calculation of variances.[10] Sensitivity and specificity were calculated for each study and for all reported cutoff values. For asymptomatic women, we plotted sensitivities and specificities in receiver operating characteristic (ROC) plots according to the timing of transvaginal ultrasonography (20-24, and >24 weeks of gestation) and definition of spontaneous preterm birth as outcome measure (<28, <32, <34, and <37 weeks of gestation). We then constructed summary ROC curves for each outcome using a bivariate random-effects approach[11] and calculated area under the summary ROC curves with their corresponding 95% confidence intervals (CIs).[12] This measure allows for comparison of the predictive accuracy of the test for different outcomes and cutoff values (two-sided P < .05 was considered to be statistically significant).

Metaanalyses were performed using subgroups of studies with a similar gestational age at testing, cutoff values, and outcome measures to minimize clinical heterogeneity. Pooled estimates of sensitivity and specificity with 95% CIs were calculated using bivariate, random-effects meta-regression model.[11] Thereafter, we derived likelihood ratios with 95% CIs from the pooled sensitivities and specificities for each outcome reported.[13] Likelihood ratios indicate by how much a given test result raises or lowers the probability of having the disease and thus allow interpretation of the results for use in clinical practice.[14] Likelihood ratios for a positive test result >10 and likelihood ratios for a negative test result <0.1 have been noted as providing convincing predictive evidence. Moderate prediction can be achieved with likelihood ratio values of 5-10 and 0.1-0.2, whereas those <5 and > 0.2 would provide only minimal prediction.[14] Likelihood ratios are more clinically meaningful than sensitivities or specificities because they are less likely to change with the prevalence of the disorder, they can be calculated for several levels of the test, and they can also be used in conjunction with pretest probability of disease to estimate the post-test probability of disease for individual patients.

Likelihood ratios generated from metaanalyses were used to determine post-test probabilities of spontaneous preterm birth <28, <32, <34, and <37 weeks of gestation for positive and negative CL results as follows:[14]

Posttest probability of preterm birth = likelihood ratio × pretest probability/[1-pretest probability × (1-likelihood ratio)]

Estimates of pretest probabilities of spontaneous preterm birth <28, <32, <34, and <37 weeks of gestation were obtained from the global prevalence of these outcomes among included studies.

Heterogeneity of the results among studies was investigated through visual examination of forest plots of sensitivities and specificities, and ROC plots. In addition, heterogeneity was assessed by means of the quantity I^2, which describes the percentage of total variation across studies that is due to heterogeneity rather than chance.[15] Statistical heterogeneity was defined as an I^2 statistic value of ≥50%.[15] We explored potential sources of heterogeneity by performing meta-regression analysis of subgroups as defined a priori:[16] study setting (those conducted in North America *vs* Europe), sample size (<100 *versus* ≥100 in studies of asymptomatic women and <50 *versus* ≥50 in studies of symptomatic women), and study's year of publication (<2000 vs ≥ 2000). In addition, we examined the impact of study quality on estimation of predictive accuracy according to individual quality items and also according to an overall quality level incorporating these items (those that met all 4 methodological criteria vs <4).

We assessed publication and related biases visually by examining the symmetry of funnel plots and statistically by using the Egger's regression test.[17] P<.1 indicated significant asymmetry.

The bivariate models were fitted using the NLMIXED procedure (SAS 9.1 for Windows [SAS Institute, Inc., Cary, NC]). The summary ROC curves were constructed using Review Manager 5.0.21 (The Nordic Cochrane Centre, Copenhagen, Denmark). The remaining analyses were performed using SPSS version 15.0 (SPSS Inc, Chicago, IL).

RESULTS

The searches produced 1027 citations of which 314 were considered relevant (Figure 1). In all, 293 studies were excluded, the main reasons being the inclusion of only singleton pregnancies (37%), not a test accuracy study (27%), and the lack of original data (24%). A total of 21 studies, including 3523 women with twin pregnancies,[18-38] met the inclusion criteria of which 15 (3001 women) provided data for metaanalyses. Sixteen studies (3213 women) provided data on asymptomatic women[18-33] and 5 studies (310 women) on women with symptoms of preterm labor.[34-38] One of the included articles[30] was an extended analysis of an initial study[39] from which we extracted data for metaanalyses of CL ≤25 and ≤35 mm at 20-24 weeks of gestation to predict preterm birth <28 weeks of gestation. Six studies[19-21,28,32,37] could not be included in the metaanalyses because the CL cutoff values used,[19-21,32] and outcome measures evaluated[28,37] did not match those of other studies that provided data for metaanalyses. Predictive accuracy of CL for spontaneous preterm birth reported in studies not included in the metaanalyses is provided in the supplementary Appendix (available online at www.ajog.org).

The main characteristics of studies included in the review are shown in Table 1. Nine studies (43%) were performed in Europe, 8 (38%) in the United States, 2 in Israel, and 1 each in Egypt and the French West Indies. The sample size ranged from 18[28] to 1135[30] in asymptomatic women (median, 139) and from 26[34,36] to 105[35] in symptomatic women (median, 66). The range for outcome definitions was preterm birth at <28 to <37 weeks of gestation. Three studies provided data on preterm birth <28 weeks, 7 on preterm birth <32 weeks, 16 on preterm birth <34 or <35 weeks, and 8 on preterm birth <36 or <37 weeks. Five studies reported results using a cervical length cutoff value of 20 mm, 13 using 25 mm, 9 using 30 mm, and 8 using 35 mm. Among asymptomatic women, 3 studies reported data on CL at <20 weeks, 12 on CL at 20-24 weeks, and 8 on CL at >24 weeks.

Methodological quality of studies included in the systematic review is summarized in Figure 2. Six studies (29%), 5 among asymptomatic and 1 among symptomatic women, met all 4 criteria; and 9 studies (43%), 7 among asymptomatic and 2 among symptomatic women, met 3 criteria (Figure 2). The most common shortcoming was failure to blind investigators to CL results. In 7 (33%) studies, results of CL were made available to the physicians involved in the clinical treatment of the patients, whereas in 8 (38%) studies it was unclear whether CL results were provided to caregivers.

Figure 3 shows the summary ROC curves of CL for the prediction of spontaneous preterm birth in asymptomatic women tested at 20-24 weeks of gestation. The greatest area under the summary ROC curve was for a CL ≤25 mm to predict preterm birth <28 weeks (0.86) (Figure 3a), followed by CL ≤20 and ≤25 mm to predict preterm birth <32 weeks (both 0.80) (Figure 3b), and cervical length ≤20 and ≤35 mm to predict preterm birth <34 weeks (both 0.77) (Figure 3c). The remaining areas under the summary ROC curves ranged between 0.64 and 0.76 (Figure 3). Among asymptomatic women tested > 24 weeks of gestation, the areas under the summary ROC curves for a cervical length ≤25 mm to predict preterm birth <32, 34, and 37 weeks of gestation were 0.77, 0.62 and 0.60, respectively (figures not shown).

Pooled estimates of accuracy of CL for the prediction of spontaneous preterm birth in women with twin pregnancies are presented in Table 2. Among asymptomatic women, pooled sensitivities and specificities ranged from 21% to 82% and from 58% to 97%, respectively. Among women with symptoms of preterm labor, pooled sensitivities and specificities varied from 49% to 79% and from 32% to 74%, respectively. Preterm birth <32 and <34 weeks of gestation in asymptomatic women were best predicted at 20-24 weeks of gestation by a CL ≤20 mm (pooled positive likelihood ratios of 10.1 and 9.0; pooled negative likelihood ratios of 0.64 and 0.74). Similar pooled estimates of accuracy were obtained for CL ≤25 mm to predict preterm birth <28 weeks of gestation (positive and negative likelihood ratios of 9.6 and 0.40, respectively). CL was less accurate to predict preterm birth <37 weeks of gestation (positive and negative likelihood ratios between 1.5 and 4.4, and between 0.71 and 0.83, respectively). A CL ≤25 mm at > 24 weeks of gestation had pooled positive and negative likelihood ratios of 1.8-2.7 and 0.47-0.75, respectively, for predicting preterm birth from <32 to <37 weeks of gestation. Among women with threatened preterm labor, the measurement of CL had a minimal predictive accuracy for preterm birth <34 and <37 weeks of gestation (pooled positive and negative likelihood ratios between 1.2 and 1.9, and between 0.65 and 0.69, respectively). It was not possible to perform metaanalysis of studies reporting data on cervical length at <20 weeks of gestation because of the differences in cutoff values and outcome measures used in the 3 studies[23,26,28] (217 women) that met the minimal inclusion criteria.

Table 3 summarizes pooled estimates of pretest and post-test probabilities of having spontaneous preterm birth in women with twin pregnancies after measurement of CL. In asymptomatic women, a CL ≤20 mm at 20-24 weeks of gestation increased the pretest probability of preterm birth <32 and <34 weeks of gestation from 6.8-42.4%, and from 15.3-61.9%, respectively, whereas a CL >20 mm decreased the risk to 4.5% and 11.8%, respectively. A cervical length ≤25 mm at 20-24 weeks of gestation increased the pretest probability of preterm birth <28 and <37 weeks of gestation from 3.5%-25.8%, and from 41.2%-75.5%, respectively, whereas a CL >25 mm decreased the risk to 1.4% and 36.8%, respectively. In both asymptomatic women tested > 24 weeks of gestation and women with symptoms of threatened labor, the likelihood ratios only produced minimal changes in the pretest probabilities of preterm birth.

There was graphical and statistical heterogeneity of predictive performance among studies as confirmed by I^2 values > 50% in some of the meta-analyses performed. An explanation for heterogeneity was not provided by the study setting, sample size, and study's year of publication. Most women included in the current metaanalyses were enrolled in a single study.[30] Sensitivity analyses, in which we removed this study, revealed similar pooled accuracy results (data not shown). Individual quality items did not have effect on predictive ability of CL at 20-24 weeks of gestation. In addition, pooled predictive accuracy estimates obtained from studies that met <4 methodological criteria did not differ significantly from those obtained from studies that met all 4 criteria. All funnel plots showed no asymmetry, either visually or in terms of statistical significance (P>.10 for all, by Egger test), indicating that publication and related biases were not present.

COMMENT

This systematic review and metaanalysis gives the strongest evidence to date that transvaginal sonographic measurement of CL at 20-24 weeks of gestation is a good predictor of spontaneous preterm birth in asymptomatic women with twin pregnancies. A CL ≤20 mm predicts spontaneous preterm birth at <32 and <34 weeks of gestation, whereas a CL ≤25 mm predicts preterm birth at <28 weeks of gestation. A "normal" CL, however, was less accurate in predicting the absence of preterm birth because the likelihood ratios for negative test results generated only minimal changes in the pretest probabilities of preterm birth. In addition, transvaginal sonographic CL has limited accuracy in predicting spontaneous preterm birth in women with twin pregnancies and threatened preterm labor, and in asymptomatic women in which the test is performed >24 weeks of gestation. There were limited data on the predictive accuracy of cervical length performed <20 weeks of gestation.

The strength of our review is based upon its compliance with stringent criteria for performing a rigorous systematic review of predictive test accuracy. These included, among others, the use of a prospective protocol designed to address a research question; the extensive search of relevant studies without language restrictions; the use of well developed methods for quality assessment and techniques recently recommended for metaanalysis of diagnostic and predictive tests, and the investigation for possible sources of heterogeneity. However, some potential limitations of our study must also be considered. First, like any systematic review, it is limited by the quality of included studies. The main area where quality was poor was in the area of blinding of the results of the sonographic CL measurement. There is evidence to suggest that this bias can lead to a significant overestimation of predictive accuracy.[7,40,41] Nevertheless, both subgroup and meta-regression analyses did not reveal the blinding of test results to significantly affect predictive performance. Second, there was significant heterogeneity among individual studies in some of the metaanalyses performed. Homogeneity is one of the desired prerequisites for metaanalysis, but it is not an absolute requirement. We explored the sources of heterogeneity as thoroughly as possible but we were unable to explain it. In the presence of unexplained heterogeneity, the use of a random-effects meta-regression model, which we did, provides the most useful estimate for informing practice. Third, the statistical power of some of our metaanalyses was limited by the small number of studies within each subgroup and the relatively small sample size of some included studies.

Only 1 previous systematic review, which included 14 studies involving 1593 women, has evaluated the accuracy of transvaginal sonographic CL in predicting spontaneous preterm birth in twin pregnancies.[4] In 2003, Honest et al[4] reported pooled positive and negative likelihood ratios between 1.5 and 5.0 and between 0.56 and 1.17, respectively, to predict spontaneous preterm birth from <32 to <37 weeks of gestation in asymptomatic women by using cutoff values for CL ranging between 20-35 mm at 20-24 weeks of gestation. Pooled

positive and negative likelihood ratios for predicting preterm birth at <34 weeks of gestation in asymptomatic women using cutoff values for CL ranging between 25-35 mm at >24 weeks of gestation varied between 1.8 and 2.1 and between 0.29 and 0.83, respectively. This review, however, included 2 studies with duplicate data and one study in which a cervical cerclage was placed in women with a CL <30 mm at <27 weeks of gestation.

The results of our systematic review suggest that transvaginal sonographic measurement of CL is a better predictor of spontaneous preterm birth in twin pregnancies than in singleton pregnancies. In fact, the metaanalysis by Honest et al[4] reported that among asymptomatic women with singleton pregnancies, a CL ≤25 mm at 20-24 weeks of gestation had pooled positive and negative likelihood ratios of 4.2 and 0.40, and 4.4 and 0.67, respectively, to predict preterm birth <32 and <34 weeks of gestation, respectively. At <20 weeks of gestation, a CL ≤25 mm had pooled positive and negative likelihood ratios of 4.1 and 0.75, and 6.3 and 0.79, respectively, to predict preterm birth <32 and <34 weeks of gestation, respectively, whereas at >24 weeks of gestation the pooled positive and negative likelihood ratios for predicting preterm birth at <34 weeks of gestation were 4.1 and 0.62, respectively. In contrast, the present study showed that among asymptomatic women with twin pregnancies, a CL ≤20 mm at 20-24 weeks of gestation had pooled positive and negative likelihood ratios of 10.1 and 0.64, and 9.0 and 0.64, respectively, to predict preterm birth at <32 and <34 weeks of gestation, respectively.

Although our metaanalyses have demonstrated that transvaginal sonographic measurement of CL is predictive of preterm birth in asymptomatic women with twin pregnancies, it is unclear if antenatal management of these pregnancies based on the results of this test can prevent preterm birth. In 2006, Gordon et al[42] conducted a study in which 125 women with twin pregnancies were randomly assigned to undergo a transvaginal sonographic CL measurement and a cervical digital exam every 4 weeks starting at 16-20 weeks until 28 weeks gestation (n=63) or to a digital cervical exam without cervical assessment by ultrasound at the same intervals (n=62). Women who underwent transvaginal sonographic cervical examinations were treated with a predetermined algorithm for the use of cerclage and bedrest. Treatment decisions of women allocated to the control group were not based upon a predetermined algorithm. There was no significant difference between groups in mean length of gestation (35.7±2.2 weeks in the study group *versus* 35.5±3.1 weeks in the control group, P=0.60) but life table analysis revealed that preterm birth <35 weeks of gestation in the transvaginal sonographic CL group was significantly reduced (P=0.02). In addition, it should be emphasized that as of late 2009, there is no therapy that effectively prevents preterm birth in twin pregnancies. This includes the use of bed rest,[43] oral betamimetics,[44] cervical cerclage,[45] and progesterone.[46]

Further well-designed randomized controlled trials are required to evaluate the effectiveness of antenatal management of women with twin pregnancies based upon transvaginal sonographic CL results for the prevention of preterm birth. In addition, these studies should include a clear protocol for the management of women based on the results of the test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Martin JA, Hamilton BE, Sutton PD, Ventura SJ, Menacker F, Kirmeyer S, Mathews TJ. Births: final data for 2006. Natl Vital Stat Rep. 2009; 57:1–102.

2. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet. 2008; 371:75–84. [PubMed: 18177778]

3. 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; 181:1465–72. [PubMed: 10601930]

4. 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; 22:305–22. [PubMed: 12942506]

5. Crane JM, Hutchens D. Transvaginal sonographic measurement of cervical length to predict preterm birth in asymptomatic women at increased risk: a systematic review. Ultrasound Obstet Gynecol. 2008; 31:579–87. [PubMed: 18412093]

6. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med. 2008; 149:889–97. [PubMed: 19075208]

7. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JHP, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA. 1999; 282:1061–6. [PubMed: 10493205]

8. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003; 3:25. [PubMed: 14606960]

9. Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med Res Methodol. 2005; 5:19. [PubMed: 15918898]

10. Sankey S, Weisfiels L, Fine M, Kapoor W. An assessment of the use of the continuity correction for sparse data in meta-analysis. Commun Stat Simulation Computation. 1996; 25:1031–56.

11. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005; 58:982–90. [PubMed: 16168343]

12. Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. Stat Med. 2002; 21:1237–56. [PubMed: 12111876]

13. Zwinderman AH, Bossuyt PM. We should not pool diagnostic likelihood ratios in systematic reviews. Stat Med. 2008; 27:687–97. [PubMed: 17611957]

14. 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; 271:703–7. [PubMed: 8309035]

15. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003; 327:557–60. [PubMed: 12958120]

16. Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat Med. 2002; 21:1525–37. [PubMed: 12111918]

17. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analyses detected by a simple graphical test. BMJ. 1997; 315:629–34. [PubMed: 9310563]

18. Goldenberg RL, Iams JD, Miodovnik M, Van Dorsten JP, Thurnau G, Bottoms S, Mercer BM, Meis PJ, Moawad AH, Das A, Caritis SN, McNellis D. The preterm prediction study: risk factors in twin gestations. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. Am J Obstet Gynecol. 1996; 175:1047–53. [PubMed: 8885774]

19. Imseis HM, Albert TA, Iams JD. Identifying twin gestations at low risk for preterm birth with a transvaginal sonographic cervical measurement at 24 to 26 weeks' gestation. Am J Obstet Gynecol. 1997; 177:1149–55. [PubMed: 9396911]

20. Wennerholm UB, Holm B, Mattsby-Baltzer I, Nielsen T, Platz-Christensen J, Sundell G, Hosseini N, Hagberg H. Fetal fibronectin, endotoxin, bacterial vaginosis and cervical length as predictors of preterm birth and neonatal morbidity in twin pregnancies. Br J Obstet Gynaecol. 1997; 104:1398–404. [PubMed: 9422019]

21. Grisaru-Granovsky S, Farine D, Barrett J, Van Eyk N, Ryan G, Seaward PGR, Windrim R. Is a single ultrasound measurement of cervical length a predictor of the risk of preterm delivery in multifetal pregnancy? Am J Obstet Gynecol. 1998; 178(1S):191S.

22. Yang JH, Kuhlman K, Daly S, Berghella V. Prediction of preterm birth by second trimester cervical sonography in twin pregnancies. Ultrasound Obstet Gynecol. 2000; 15:288–91. [PubMed: 10895446]

23. Guzman ER, Walters C, O'reilly-Green C, Kinzler WL, Waldron R, Nigam J, Vintzileos AM. Use of cervical ultrasonography in prediction of spontaneous preterm birth in twin gestations. Am J Obstet Gynecol. 2000; 183:1103–7. [PubMed: 11084549]

24. Soriano D, Weisz B, Seidman DS, Chetrit A, Schiff E, Lipitz S, Achiron R. The role of sonographic assessment of cervical length in the prediction of preterm birth in primigravidae with twin gestation conceived after infertility treatment. Acta Obstet Gynecol Scand. 2002; 81:39–43. [PubMed: 11942885]

25. Vayssière C, Favre R, Audibert F, Chauvet MP, Gaucherand P, Tardif D, Grangé G, Novoa A, Descamps P, Perdu M, Andrini E, Janse-Marec J, Maillard F, Nisand I. Cervical length and funneling at 22 and 27 weeks to predict spontaneous birth before 32 weeks in twin pregnancies: a French prospective multicenter study. Am J Obstet Gynecol. 2002; 187:1596–604. [PubMed: 12501070]

26. Gibson JL, Macara LM, Owen P, Young D, Macauley J, Mackenzie F. Prediction of preterm delivery in twin pregnancy: a prospective, observational study of cervical length and fetal fibronectin testing. Ultrasound Obstet Gynecol. 2004; 23:561–6. [PubMed: 15170796]

27. Sperling L, Kiil C, Larsen LU, Qvist I, Bach D, Wojdemann K, Bladh A, Nikkila A, Jørgensen C, Skajaa K, Bang J, Tabor A. How to identify twins at low risk of spontaneous preterm delivery. Ultrasound Obstet Gynecol. 2005; 26:138–44. [PubMed: 16038015]

28. Fait G, Har-Toov J, Gull I, Lessing JB, Jaffa A, Wolman I. Cervical length, multifetal pregnancy reduction, and prediction of preterm birth. J Clin Ultrasound. 2005; 33:329–32. [PubMed: 16196008]

29. Arabin B, Roos C, Kollen B, van Eyck J. Comparison of transvaginal sonography in recumbent and standing maternal positions to predict spontaneous preterm birth in singleton and twin pregnancies. Ultrasound Obstet Gynecol. 2006; 27:377–86. [PubMed: 16514620]

30. To MS, Fonseca EB, Molina FS, Cacho AM, Nicolaides KH. Maternal characteristics and cervical length in the prediction of spontaneous early preterm delivery in twins. Am J Obstet Gynecol. 2006; 194:1360–5. [PubMed: 16647922]

31. Klein K, Gregor H, Hirtenlehner-Ferber K, Stammler-Safar M, Witt A, Hanslik A, Husslein P, Krampl E. Prediction of spontaneous preterm delivery in twin pregnancies by cervical length at mid-gestation. Twin Res Hum Genet. 2008; 11:552–7. [PubMed: 18828739]

32. Aboulghar MM, Aboulghar MA, Mourad L, Serour GI, Mansour RT. Ultrasound cervical measurement and prediction of spontaneous preterm birth in ICSI pregnancies: a prospective controlled study. Reprod Biomed Online. 2009; 18:296–300. [PubMed: 19192354]

33. Fox NS, Saltzman DH, Klauser CK, Peress D, Gutierrez CV, Rebarber A. Prediction of spontaneous preterm birth in asymptomatic twin pregnancies with the use of combined fetal fibronectin and cervical length. Am J Obstet Gynecol. 2009; 201:313.e1–5. [PubMed: 19733285]

34. 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; 90:357–63. [PubMed: 9277644]

35. Persutte WH, Chyu J, Cioffi-Ragan D, Hobbins JC. Cervical length in twins. Am J Obstet Gynecol. 2000; 182:S118.

36. 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; 72:117–25. [PubMed: 11166744]

37. Fuchs I, Tsoi E, Henrich W, Dudenhausen JW, Nicolaides KH. Sonographic measurement of cervical length in twin pregnancies in threatened preterm labor. Ultrasound Obstet Gynecol. 2004; 23:42–5. [PubMed: 14970998]

38. Gonzalez N, Bige V, Kandoussi S, Graesslin O, Quereux C, Gabriel R. Sonographic measurement of cervical length in twin pregnancies with preterm labor: comparison with singleton pregnancies [Article in French]. Gynecol Obstet Fertil. 2004; 32:122–7. [PubMed: 15123134]

39. Souka AP, Heath V, Flint S, Sevastopoulou I, Nicolaides KH. Cervical length at 23 weeks in twins in predicting spontaneous preterm delivery. Obstet Gynecol. 1999; 94:450–4. [PubMed: 10472876]

40. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med. 2004; 140:189–202. [PubMed: 14757617]

41. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ. 2006; 174:469–76. [PubMed: 16477057]

42. Gordon M, Robbins A, McKenna D, Howard B, Barth W. Cervical length assessment as a resource to identify twins at risk for preterm delivery (Clarity study). Am J Obstet Gynecol. 2006; 195:S55.

43. Crowther CA. Hospitalisation and bed rest for multiple pregnancy. Cochrane Database Syst Rev. 2001; 1 CD000110.

44. Yamasmit W, Chaithongwongwatthana S, Tolosa JE, Limpongsanurak S, Pereira L, Lumbiganon P. Prophylactic oral betamimetics for reducing preterm birth in women with a twin pregnancy. Cochrane Database Syst Rev. 2005; 3 CD004733.

45. Jorgensen AL, Alfirevic Z, Tudur Smith C, Williamson PR. cerclage IPD Meta-analysis Group. Cervical stitch (cerclage) for preventing pregnancy loss: individual patient data meta-analysis. BJOG. 2007; 114:1460–76. [PubMed: 17903224]

46. Norman JE, Mackenzie F, Owen P, Mactier H, Hanretty K, Cooper S, Calder A, Mires G, Danielian P, Sturgiss S, MacLennan G, Tydeman G, Thornton S, Martin B, Thornton JG, Neilson JP, Norrie J. Progesterone for the prevention of preterm birth in twin pregnancy (STOPPIT): a randomised, double-blind, placebo-controlled study and meta-analysis. Lancet. 2009; 373:2034–40. [PubMed: 19523680]
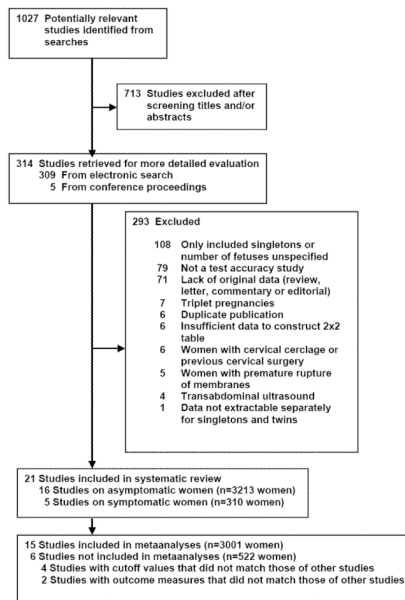
1027 Potentially relevant studies identified from searches

713 Studies excluded after screening titles and/or abstracts

314 Studies retrieved for more detailed evaluation
309 From electronic search
5 From conference proceedings

293 Excluded

108 Only included singletons or number of fetuses unspecified
79 Not a test accuracy study
71 Lack of original data (review, letter, commentary or editorial)
7 Triplet pregnancies
6 Duplicate publication
6 Insufficient data to construct 2x2 table
6 Women with cervical cerclage or previous cervical surgery
5 Women with premature rupture of membranes
4 Transabdominal ultrasound
1 Data not extractable separately for singletons and twins

21 Studies included in systematic review
16 Studies on asymptomatic women (n=3213 women)
5 Studies on symptomatic women (n=310 women)

15 Studies included in metaanalyses (n=3001 women)
6 Studies not included in metaanalyses (n=522 women)
4 Studies with cutoff values that did not match those of other studies
2 Studies with outcome measures that did not match those of other studies

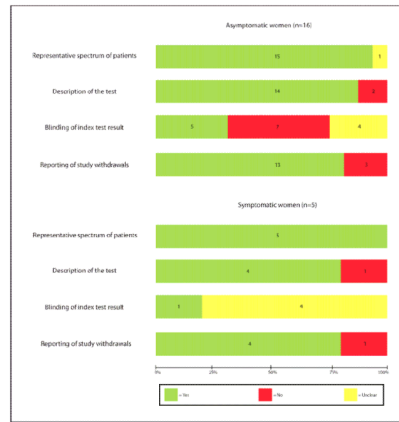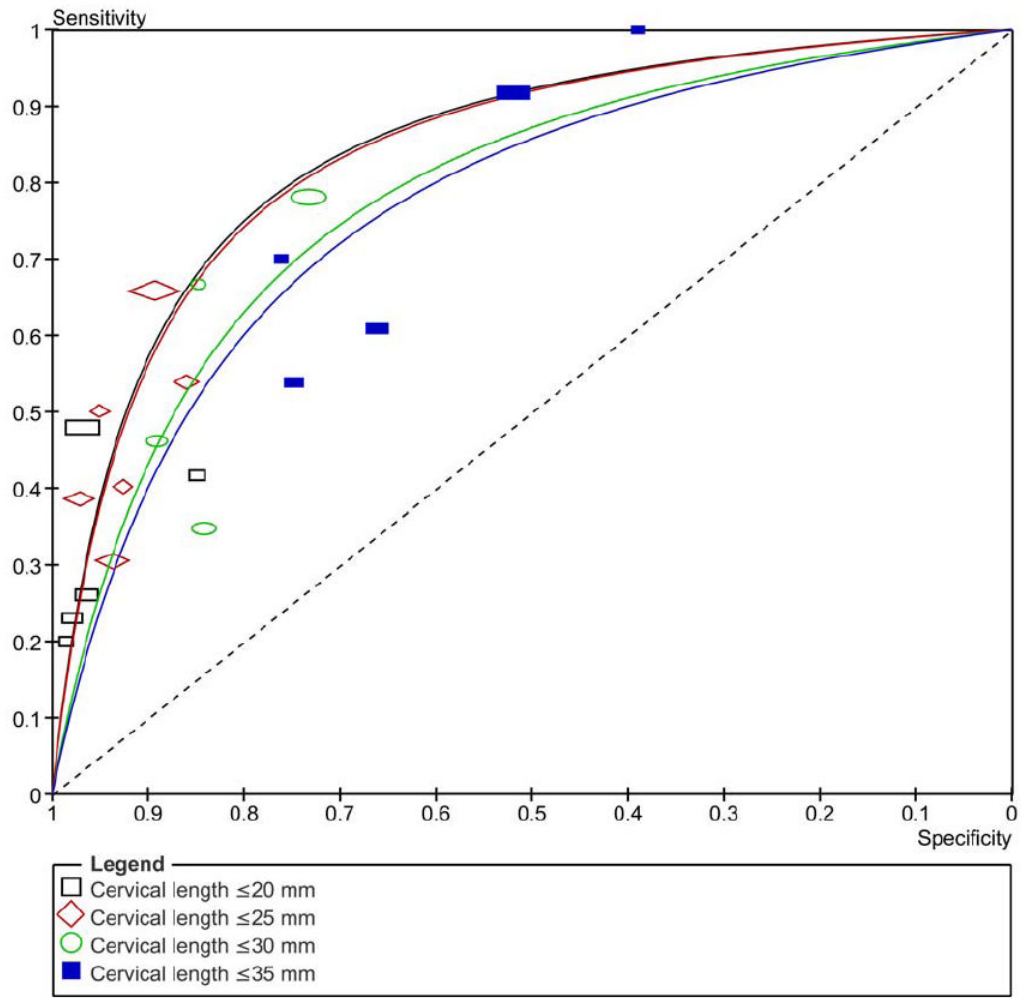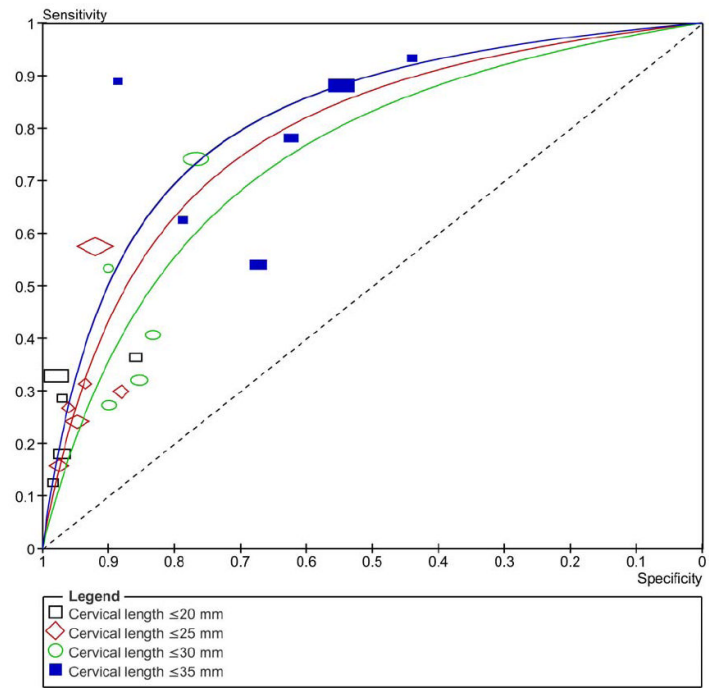**Figure 1.**
Study selection process

Figure 2.
Methodological quality of studies included in the systematic review. Data presented as percentages across all included studies. Figures in the stacks represent number of studies.
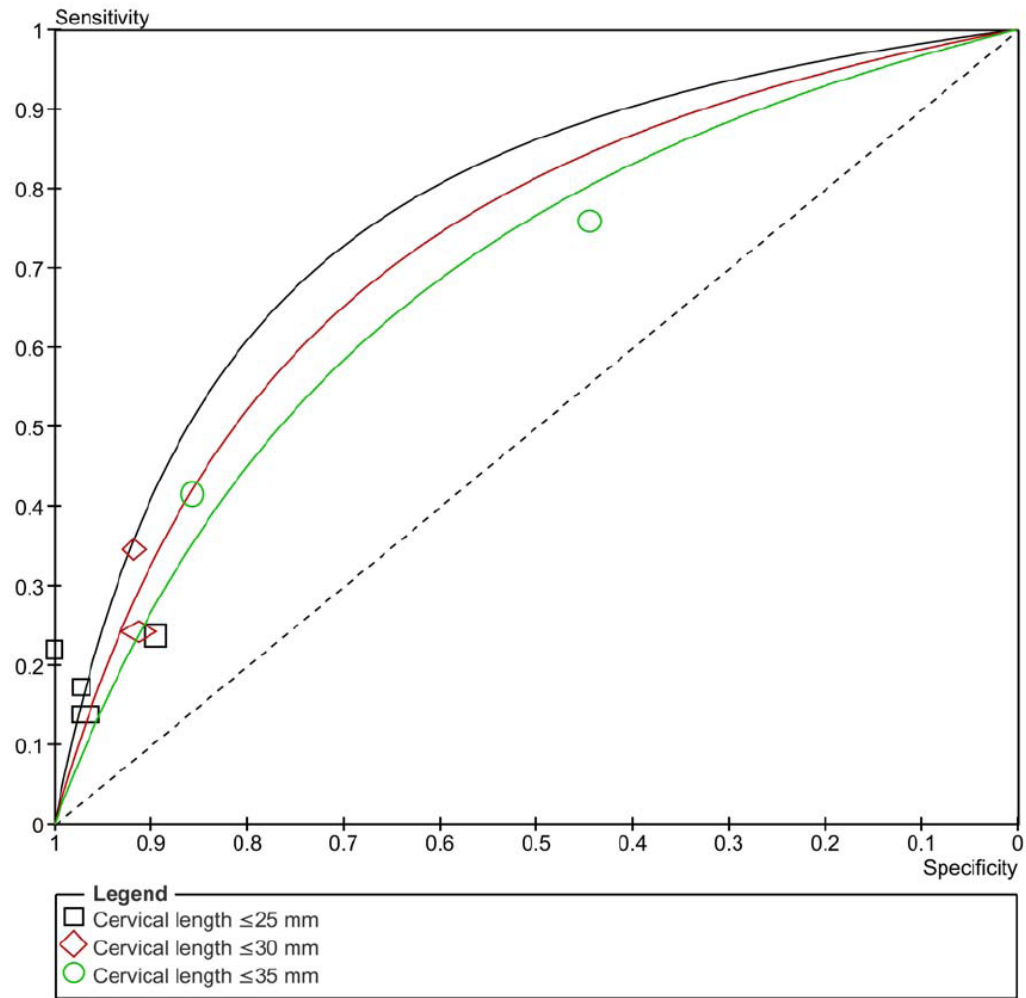
Figure 3.
Summary receiver operating characteristic (ROC) curves of cervical length tested at 20-24 weeks of gestation in asymptomatic women to predict spontaneous preterm birth: A) <28; B) <32; C) <34; D) <37 weeks of gestation. The area of each circle, rectangle and diamond is proportional to study's sample size.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

TABLE 1

Characteristics of studies included in the systematic review

Study, year,	Country	No of women	Inclusion criteria	Exclusion criteria	Gestational age at testing (weeks)	Cervical length cut-off (mm)	Outcome
Asymptomatic women							
Goldenberg,[18] 1996	United States	147	Twins	Cervical cerclage, placenta previa, major fetal anomaly	24, 28	25	Preterm birth <32, <35 and <37 weeks
Imseis,[19] 1997	United States	85	Twins	Cervical cerclage	24-26	35	Preterm birth <34 weeks
Wennerholm,[20] 1997	Sweden	101	Twins	Not reported	24-34	33	Preterm birth <35 and <37 weeks
Grisaru-Granovsky,[21] 1998	Canada	38	Twins, triplets, quadruplets	Not reported	18-29 (mean, 25)	30	Preterm birth <34 weeks
Yang,[22] 2000	United States	65	Twins	Cervical cerclage, placenta previa or bleeding	18-26 (91% at <24 weeks)	25, 30, 35	Preterm birth <32 and <35 weeks
Guzman,[23] 2000	United States	131	Twins	Cervical cerclage	15-20, 21-24, 25-28	20	Preterm birth <28,<30, <32, and <34 weeks
Soriano,[24] 2002	Israel	44	Twins	Not reported	18-24 (mean, 22.7)	35	Preterm birth <34 weeks
Vayssiere,[25] 2002	France	251	Twins	Cervical cerclage, placenta previa, major fetal anomaly, twin-twin transfusion syndrome, premature rupture of membranes	21-23, 26-28	25, 30	Preterm birth <32 and <35 weeks
Gibson,[26] 2004	United Kingdom	91	Twins	fetal anomaly, twin-twin transfusion syndrome	18,24,28, and 32	25, 22	Preterm birth <35 weeks
Sperling,[27] 2005	Denmark and Sweden	383	Twins	Cervical cerclage, prior conization	23	20, 25, 30, 35	Preterm birth <28,<32, <33, <34 and <35 weeks
Fait,[28] 2005	Israel	18	Triplets reduced to twins	Cervical cerclage	15.9 ± 0.3	35	Preterm birth <33 weeks
Arabin,[29] 2006	The Netherlands	153	Twins	Not reported	20-25	25, 30	Preterm birth <36 weeks

Study, year,	Country	No of women	Inclusion criteria	Exclusion criteria	Gestational age at testing (weeks)	Cervical length cut-off (mm)	Outcome
To,[30] 2006	United Kingdom	1135	Twins	Cervical cerclage, major fetal abnormalities, premature rupture of membranes, monochorionic twins with severe twin–twin transfusion syndrome	22-24	5-55[a]	Preterm birth <30, <32, and <34 weeks
Klein,[31] 2008	Austria	223	Twins	Not reported	20-25	25, 30, 35	Preterm birth <34 weeks
Aboulghar,[32] 2009	Egypt	193	Twin ICSI pregnancies	Cervical cerclage	Mean, 20	38	Preterm birth <34 weeks
Fox,[33] 2009	United States	155	Twins	Monoamniotic twins, fetal aneuploidy, major fetal abnormalities	22-24, 25-32	20, 25, 35	Preterm birth <28,<32, <34,and <37 weeks
Women with symptoms of preterm labor							
Crane,[34] 1997	Canada	26	Twins, regular uterine contractions with cervical changes (dilatation, effacement, or change in consistency)	Active vaginal bleeding, placenta previa, premature rupture of membranes, cervical cerclage, stillbirth, cervical dilatation >3 cm	23-33	25, 30	Preterm birth <34 and <37 weeks
Persutte,[35] 2000	United States	105	Twins, spontaneous onset of labor and vaginal delivery	Not reported	20-32	25	Preterm birth <37 weeks
Vendittelli,[36] 2001	French West Indies	26	Twins, uterine contractions on external tocodynamometry at least one every 10 min and/or with changes in manual cervical effacement or dilatation	Cervical dilatation >3 cm, premature rupture of membranes, cervical cerclage, active vaginal bleeding, placenta previa, stillbirth, fetal malformation,	18-36	25	Preterm birth <37 weeks
Fuchs,[37] 2004	Germany, United Kingdom	87	Twins, painful and regular uterine contractions	Cervical dilatation >3 cm, cervical cerclage, premature rupture of membranes	24-36	25	Delivery within 7 days of testing
Gonzalez,[38] 2004	France	66	Twins, regular uterine contractions with cervical changes (effacement of at least 50% or dilatation of at least one finger)	Cervical cerclage, maternal or fetal disorder contraindicating continuation of pregnancy	<34	20, 30	Preterm birth <34 and <37 weeks

[a] For the review, we chose <20, <25, <30, and <35 mm as cutoff values

TABLE 2

Pooled estimates for cervical length in predicting spontaneous preterm birth in women with twin pregnancies

Outcome	Cervical length cut-off (mm)	No of studies	Sample size	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Positive Likelihood ratio (95% CI)	Negative Likelihood ratio (95% CI)
Asymptomatic women							
Testing at 20-24 weeks' gestation							
Preterm birth <28 weeks	20	3 [23,27,33]	591	35 (14-62)	93 (91-95)	5.2 (2.6-10.6)	0.69 (0.49-1.01)
	25	3 [27,33,39]	637	64 (41-83)	93 (91-95)	9.6 (5.8-14.8)	0.40 (0.23-0.68)
	35	3 [27,33,39]	637	82 (60-95)	66 (62-69)	2.4 (1.9-3.0)	0.28 (0.11-0.67)
Preterm birth <32 weeks	20	5 [23,25,27,30,33]	1955	39 (31-48)	96 (95-97)	10.1 (7.4-13.9)	0.64 (0.55-0.73)
	25	6 [18,22,25,27,30,33]	2036	54 (45-62)	91 (90-92)	6.0 (4.8-7.4)	0.51 (0.43-0.61)
	30	4 [22,25,27,30]	1812	65 (56-74)	78 (76-80)	3.0 (2.5-3.5)	0.45 (0.35-0.57)
	35	5 [22,25,27,30,33]	1889	81 (73-87)	58 (56-61)	1.9 (1.7-2.2)	0.33 (0.23-0.48)
Preterm birth <34 weeks	20	5 [23,26,27,30,33]	1760	29 (23-35)	97 (96-98)	9.0 (6.1-12.7)	0.74 (0.68-0.80)
	25	6 [18,22,27,30,31,33]	1987	40 (38-46)	93 (92-94)	5.8 (4.5-7.2)	0.64 (0.58-0.71)
	30	5 [22,25,27,30,31]	2014	56 (50-62)	81 (79-83)	3.0 (2.6-3.4)	0.55 (0.48-0.63)
	35	6 [22,24,27,30,31,33]	1884	79 (74-84)	60 (57-62)	2.0 (1.8-2.2)	0.35 (0.27-0.44)
Preterm birth <37 weeks	25	4 [18,22,29,33]	434	21 (15-27)	95 (92-98)	4.4 (2.4-8.2)	0.83 (0.75-0.92)
	30	2 [22,29]	218	29 (18-43)	91 (86-95)	3.4 (1.6-6.7)	0.78 (0.65-0.92)
	35	2 [22,33]	134	56 (43-68)	63 (50-74)	1.5 (1.0-2.2)	0.71 (0.51-0.98)
Testing after 24 weeks' gestation							
Preterm birth <32 weeks	25	3 [18,25,33]	511	65 (45-81)	76 (72-79)	2.7 (2.0-3.6)	0.47 (0.29-0.76)
Preterm birth <34 weeks	25	4 [18,25,26,33]	594	44 (34-53)	81 (78-85)	2.3 (1.8-3.1)	0.70 (0.59-0.83)
Preterm birth <37 weeks	25	2 [18,33]	276	43 (35-51)	77 (68-84)	1.8 (1.3-2.6)	0.75 (0.63-0.89)
Women with symptoms of preterm labor							
Preterm birth <34 weeks	30	2 [34,38]	92	79 (62-91)	32 (20-45)	1.2 (0.9-1.7)	0.67 (0.31-1.44)
Preterm birth <37 weeks	25	2 [34,35]	131	49 (38-60)	74 (59-87)	1.9 (1.1-3.3)	0.69 (0.52-0.90)
	30	3 [34,36,38]	118	76 (65-84)	37 (19-58)	1.2 (0.9-1.7)	0.65 (0.29-1.31)

[a]Study # 39 is an initial report of study # 30

TABLE 3

Pooled estimates of pretest probabilities, likelihood ratios, and post-test probabilities for cervical length in the prediction of spontaneous preterm birth in women with twin pregnancies

Outcome	Cervical length cut-off (mm)	Pretest probability (%)	Likelihood ratios		Post-test probabilities (%)	
			Positive test result	Negative test result	Positive test result	Negative test result
Asymptomatic women (testing at 20-24 weeks' gestation)						
Preterm birth <28 weeks	25	3.5	9.6	0.40	25.8	1.4
Preterm birth <32 weeks	20	6.8	10.1	0.64	42.4	4.5
Preterm birth <34 weeks	20	15.3	9.0	0.74	61.9	11.8
Preterm birth <37 weeks	25	41.2	4.4	0.83	75.5	36.8
Asymptomatic women (testing after 24 weeks' gestation)						
Preterm birth <32 weeks	25	6.1	2.7	0.47	14.9	3.0
Preterm birth <34 weeks	25	18.2	2.3	0.70	33.9	13.5
Preterm birth <37 weeks	25	53.6	1.8	0.75	67.5	46.4
Women with symptoms of threatened labor						
Preterm birth <34 weeks	30	35.9	1.2	0.67	40.2	27.3
Preterm birth <37 weeks	25	67.2	1.9	0.69	79.6	58.6
	30	67.2	1.2	0.65	71.1	57.1