Background Preterm birth may result in short and long-term health problems for the child. Accurate diagnosis and exclusion of preterm birth could ensure appropriate admissions into hospital, transfers to specialist units and discharge home. A systematic review was conducted to assess the diagnostic accuracy of three diagnostic tests for preterm labour in symptomatic women with intact membranes.

Methods Nine bibliographic databases were searched up to July 2017 to identify studies assessing the diagnostic accuracy of fetal fibronectin (fFN, ≥50 ng/ml), Actim Partus (AP) or PartoSure (PS) for predicting preterm birth within 7 days in women with intact membranes. Extensive supplementary searches were also conducted. Standardised data extraction and quality assessment using QUADAS-2 were undertaken. Narrative synthesis was used. All systematic review tasks were performed by at least 2 independent reviewers.

Results 2619 unique records were retrieved, of which 442 proceeded to full-text screening and 31 records (20 studies, 16 AP, 4 PS and 2 fFN – some studies looked at >1 test) were included in the systematic review. QUADAS-2 assessment found only 5 out of 20 studies to be at low risk of bias. For AP, the best overall sensitivity and specificity results were 94.7% (95% CI 89.9 to 97.7) and 92.4% (88.9–95.1), while the worst were 33.3% (4.3–77.7) and 74.1% (69.1–78.6). For PS, the best overall sensitivity and specificity results were 100.0% (73.5–100.0) and 95.4% (88.6–98.7), while the worst were 0.0% (0.0–97.5) and 97.3% (96.8–99.9). For fFN at 10 ng/ml, sensitivity ranged from 93.8% (82.8–98.7) to 95.7% (87.8–99.1), and specificity from 32.2% (27.7–37.0) to 42.3% (36.5–48.4). For fFN at 200 ng/ml, sensitivity ranged from 70.8% (55.9–83.0) to 71.0% (38.8–81.3), and specificity from 78.6% (74.3–82.5) to 83.6% (78.8–87.8). For fFN at 500 ng/ml, sensitivity ranged from 29.2% (17.0–44.1) to 42.0% (30.2–54.5), and specificity from 94.3% (91.6–96.4) to 95.7% (92.7–97.8). One study compared fFN with AP and results depended on the fFN threshold used. One study compared AP and PS and found little difference in sensitivity and specificity.

Conclusion A wide range of diagnostic accuracy estimates were provided. Substantial methodological, clinical and statistical heterogeneity between studies raises considerable uncertainty about the most valid estimate of accuracy for each index test. There were few studies comparing index tests. The current results do not allow a firm conclusion as to which test is the best to use in the diagnosis of preterm labour in symptomatic women with intact membranes. Further studies comparing index tests could help to further address this question.