Biomarker tests to help diagnose preterm labour in women with intact membranes

Produced by

Peninsula Technology Assessment Group (PenTAG)
University of Exeter Medical School
South Cloisters, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU

Authors

Dr Jo Varley-Campbell, *Research Fellow*[1]
Dr Rubén Mújica-Mota, *Senior Lecturer*[1]
Dr Helen Coelho, *Research Fellow*[1]
Dr Neel Ocean, *Postdoctoral Research Associate*[1]
Dr Max Barnish, *Postdoctoral Research Associate*[1]
Mr David Packman, *Graduate Research Associate*[1]
Miss Sophie Dodman, *Research Assistant*[1]
Mr Chris Cooper, *Senior Research Fellow*[1]
Dr Tristan Snowsill, *Research Fellow*[1,2]
Dr Tracey Kay, *Consultant Obstetrician & Gynaecologist*[3]
Dr Neil Liversedge, *Consultant Obstetrician & Gynaecologist*[3]
Dr Michelle Parr, *Consultant Neonatologist*[4]
Dr Lisa Knight, *Obstetrics & Gynaecology Registrar*[3]
Prof Chris Hyde, *Professor of Public Health and Clinical Epidemiology*[1]
Prof Andrew Shennan OBE, *Professor of Obstetrics*[5]
Assoc Prof Martin Hoyle, *Associate Professor of Health Technology Assessment*[1]

[1]Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School
[2]Health Economics Group, University of Exeter Medical School
[3]Royal Devon and Exeter NHS Foundation Trust
[4]Central Manchester University Hospital NHS Foundation Trust
[5]King's College London / Guy's and St Thomas' Hospital

Correspondence to

Dr Jo Varley-Campbell
Peninsula Technology Assessment Group (PenTAG)
University of Exeter Medical School
South Cloisters
St Luke's Campus
Heavitree Road
Exeter
EX1 2LU
j.varley-campbell@exeter.ac.uk

Date completed

10/01/2018

Source of funding

This report was commissioned by the NIHR HTA Programme as project number 17/10/01.

Declared competing interests of the authors	Professor Shennan is an investigator on a number of trials/studies related to preterm birth (GSK funded NEWBORN tocolytic trial, NIHR funded PETRA and QUIDS prediction studies, Guy's and St Thomas' Charity funded EQUIPPT, preterm management study, Tommys Charity funded preterm Birth studies). All these studies include comparing PartoSure and quantitative fetal fibronectin and have been supported by free PartoSure samples from Quiagen and financial support from Hologic (fetal fibronectin) -paid to institute to cover expenses of this comparison only. Professor Shennan has given lectures to internal staff at Biomedica (Actim Partus) and Hologic (fetal fibronectin) in the last 5 years and received financial support to cover expenses only for this when travelling to the USA and Finland. Professor Shennan has an academic interest in establishing best strategies to predict preterm labour and has published widely on the topic. He has been invited to many national and international meetings to lecture on the topic (India, Australia, Italy, South Africa, Spain, Holland in 2017), expenses paid by the organisers. No industry supported fees/expenses were paid. Professor Shennan has received financial and equipment support (paid to institution only) from other industry in the last 5 years, including GSK, A1Pharma, Clinical Innovations, Perkin Elmer and Alere but unrelated to preterm birth prediction. (Alere did market Actim Partus in the UK, and Clinical Innovations have a rupture membrane test.) He is the inventor of the CRADLE Vital Signs Alert (Gates Foundation and MRC/DFID grants), and the institution receives royalty payment for high income sales only from Microlife, but not LMIC sales. He is part inventor of Desperate Deborah and the Tydeman tube, and the royalties are paid to the institution. He has no active patents. Professor Shennan's total current grant income (active grants) as CI/PI or co-applicant is currently >£20 million. Total industry support from all sources, including royalties (paid to KCL) is currently <£200,000. All other authors declare no conflicts of interest
Rider on responsibility for report	The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.
This report should be referenced as follows:	Varley-Campbell J, Mújica-Mota R, Coelho H, Ocean N, Barnish M, Packman D, Dodman S, Cooper C, Snowsill T, Kay T, Liversedge N, Parr M, Knight L, Hyde C, Shennan A, Hoyle M. Biomarker tests to help diagnose preterm labour in women with intact membranes: 2018, Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School (Report for NICE).

Any 'commercial in confidence' data provided by companies, and specified as such, will be highlighted in blue and underlined in the assessment report. Any 'academic in confidence' data provided by companies, and specified as such, will be highlighted in yellow and underlined in the assessment report.

Abstract

Background

Preterm birth may result in short and long-term health problems for the child. Accurate diagnoses of preterm birth could prevent unnecessary (or ensure appropriate) admissions into hospital or transfers to specialist units.

Objectives

The purpose of this report is to assess the test accuracy, clinical and cost-effectiveness of the diagnostic tests PartoSure, Actim Partus and quantitative fFN at thresholds ≠ 50ng/ml (qfFN) for women presenting with signs and symptoms of preterm labour relative to fFN 50 ng/ml.

Methods

Systematic reviews were conducted of the published literature for diagnostic test accuracy studies of PartoSure, Actim Partus and qfFN for predicting preterm birth, the clinical effectiveness following treatment decisions informed by test results, and economic evaluations of the tests.

A model-based economic evaluation was also conducted to extrapolate long-term outcomes from the results of the diagnostic tests. The model followed the best methodological practice from previous published evaluations including that followed by the model that informed the 2015 NICE guidelines on preterm labour diagnosis and treatment.

Results

Twenty studies were identified evaluating diagnostic test accuracy against the reference standard of delivery within 7 days and seven against the reference standard of delivery within 48hrs. Two studies assessed two of the index tests within the same population. One study demonstrated that depending on the threshold used, qfFN was more or less accurate than Actim Partus whilst the other indicated little difference between PartoSure and Actim Partus. A study assessing qfFN and PartoSure in the same population was not identified. The test accuracy results from the other included studies revealed a high level of uncertainty, primarily due to substantial methodological, clinical and statistical heterogeneity between studies.

No clinical effectiveness studies evaluating any of the three biomarker tests were identified.

One partial economic evaluation was identified for predicting preterm birth. It assessed the number needed to treat to prevent a respiratory distress syndrome case of a 'treat-all' strategy, relative to testing with qualitative qfFN.

In our de-novo base case analysis (for woman at 30 weeks' gestation) Actim Partus had lower healthcare costs and fewer quality-adjusted life-years than qfFN 50 ng/ml, reducing costs at a rate of £56,030 per QALY lost vs. qfFN 50ng ml. PartoSure is less costly than Actim Partus whilst being equally effective, but this is based on diagnostic accuracy data from a small study. No study provided data that allowed us to compare all three tests simultaneously. Testing with qfFN at 10ng/ml cost £140,267 per QALY gained relative to fFN 50ng/ml, whilst testing with qfFN at 200ng/ml and 200ng/ml resulted in lower cost savings per QALY lost relative to fFN 50ng/ml than those with Actim Partus. Similar qualitative results obtained for women presenting at different gestational ages.

Discussion and Conclusion

There is a high degree of uncertainty surrounding the test accuracy and cost-effectiveness results. We are also aware of four ongoing UK trials, two of which plan to enrol over 1,000 participants. The results of these trials may alter the findings presented here.

Registration

PROSPERO: CRD42017072696

Funding Details

This report was commissioned by the NIHR HTA Programme as project number 17/10/01.

Word Count

492 words.

Contents

List of tables

List of figures

Abbreviations

95% CI	95% Confidence Intervals
ACS	Antenatal Corticosteroids
ANS	Antenatal steroids
CL	Cervical Length
BAPM	British Association of Perinatal Medicine
BPD	Bronchopulmonary Dysplasia
DTA	Diagnostic Test Accuracy
GA	Gestational Age
ELISA	Enzyme-linked Immunosorbent Assay
fFN	Fetal Fibronectin
ICER	Incremental Cost-Effectiveness Ratio
IQR	Interquartile Range
IVH	Intraventricular Haemorrhage
(+/-) LR	(+/-) Likelihood Ratio
NICE	National Institute for Health and Care Excellence
NICU	Neonatal intensive care unit
NPV	Negative Predictive Value
OR	Odds ratio
PAMG-1	Placental Alpha Microglobulin-1
ph(IGFBP-1)	Phosphorylated (Insulin-like Growth Factor Binding Protein-1)

PPROM	Preterm Premature Rupture of Membranes
PPV	Positive Predictive Value
PTB	Preterm Birth
QALY	Quality-Adjusted Life Year
qfFN	Quantitative fetal fibronectin
RDS	Respiratory Distress Syndrome
ROC	Receiver Operating Characteristic
RR	Risk Ratio
SD	Standard Deviation
TPB	Threatened preterm birth
TPL	Threatened preterm labour
TTO	Time trade off
WHO	World Health Organisation

Glossary

Antenatal corticosteroids (ACS)	Antenatal corticosteroid therapy is administered to women when preterm delivery is anticipated, to enhance foetal lung maturation. The aim of treatment is to prevent respiratory distress syndrome and reduce mortality and morbidity for the preterm infant.
Bronchopulmonary dysplasia (BPD)	Chronic lung disease that affects premature new-borns requiring oxygen therapy. Commonly occurs secondary to respiratory distress syndrome.
Cervical length (CL)	Cervical length measurement via transvaginal ultrasound is a technique used to assess the risk of preterm delivery in high-risk women or women presenting with signs or symptoms of preterm labour. Shortening of the cervical length is correlated with higher-risk of preterm delivery.
Cervical os	Opening of the uterine cervix (anatomy). The cervical os dilates during childbirth to allow the passage of the baby.
'Comparative' study	A study design that assesses (but does not necessarily directly compare) the performance of two different diagnostic tests within the same population.
Concordance	The proportion of cases in which the result of the test agrees with the clinical outcome.
Diagnostic yield	The number of positive results divided by the number of samples
Fetal fibronectin (fFN)	Adhesion protein that binds the fetal sac to the uterine lining. After 35 weeks gestation the protein begins to degrade to prepare for delivery. Detection of fetal fibronectin in cervico-vaginal secretions earlier than 35 weeks can be used to predict onset of preterm delivery (fFN test).
Gestational Age	The number of completed weeks of pregnancy. This is usually calculated from the first day of the woman's last menstrual period,

	or alternatively from clinical examination or ultrasonography. Reported as weeks+days.
Gravidity	The number of times a woman has been pregnant
Iatrogenic delivery	A delivery that is medically initiated or accelerated, such as administration of labour-inducing drugs or delivery via caesarean section.
Incremental Cost-Effectiveness Ratio (ICER)	Term used in health economics to compare the difference is the cost and the effectiveness of two interventions/tests. ICER= $(C_1-C_0) / (E_1-E_0)$ C_1= Cost of Intervention C_2= Cost of control E_1= Effectiveness of intervention E_2= Effectiveness of control
Intraventricular Haemorrhage (IVH)	A condition associated with preterm delivery, characterised by bleeding into the ventricles of the brain. Severity is categorised by four grades: Grade 1 & 2 smaller amount of bleeding; Grades 3 & 4 more severe bleeding.
Likelihood Ratio (LR)	The likelihood of a given test result in a patient who has a preterm delivery compared to the likelihood of the same result in a patient who does not deliver preterm. Positive LR: how much more often a positive test result occurs in people who do deliver preterm compared to those who do not. $$\text{Postive LR} = \frac{\text{P(Test + ve} \mid \text{preterm)}}{\text{P(Test + ve} \mid \text{not preterm)}} = \frac{\text{Sensitivity}}{1 - \text{Specificity}}$$ Negative LR: how much less likely a negative result is in people with preterm delivery compared to those without preterm delivery. $$\text{Negative LR} = \frac{\text{P (Test} - \text{ve} \mid \text{preterm)}}{\text{P (Test} - \text{ve} \mid \text{not preterm)}} = \frac{1 - \text{Sensitivity}}{\text{Specificity}}$$

Meta-analysis	Statistical technique that combines data from various studies evaluating the same index test to calculate pooled diagnostic accuracy estimates.
Multiple gestation pregnancies	Pregnancies where the number of foetuses exceeds one.
Negative Predictive Value (NPV)	The proportion of people with a negative result that will not deliver preterm (within 48 hours or 7 days) NPV = True Negative / (True Negative + False Negative)
Parity	The number of times a woman has carried a pregnancy to a viable gestation.
Phosphorylated Insulin-like Growth Factor Binding Protein ph (IGFBP-1)	Protein produced by decidual cells that leaks into cervical secretions when delivery is imminent and can be used to predict the onset of preterm labour (Actim Partus).
Placental alpha macroglobulin-1 (PAMG-1)	PAMG-1 protein is secreted by the decidual cells into the amniotic fluid throughout pregnancy. This protein can be detected in cervicovaginal secretions when delivery is imminent (PartoSure test).
Positive Predictive Value (PPV)	The proportion of people with a positive result that will deliver preterm (within 48 hours or 7 days). PPV = True Positive / (True Positive+ False Positive)
Preterm birth/ delivery	The delivery of a live baby before 37+0 weeks gestational age. <28 weeks GA (extremely preterm) ≥28 weeks and <32 weeks GA(very preterm) ≥32 weeks and <37 weeks GA (moderate to late preterm)
Preterm Premature Rupture of Membranes (PPROM)	Premature (<37 weeks gestation) rupture of the amniotic sac surrounding the foetus before the onset of established labour. Women experiencing PPROM are at increased risk of amniotic infection and preterm delivery.

Prevalence	The proportion of women actually delivering preterm (within 48 hour or 7 days)
Quality-Adjusted Life Year	A measure of disease burden that combines length and quality of life
Reference standard	This refers to the best diagnostic test currently available, against which an index test is assessed. Due to the predictive nature of the index tests the reference standard for all included studies was preterm delivery within 48hrs or within 7 days.
Respiratory Distress Syndrome	Breathing disorder that commonly affects premature babies due to insufficient surfactant production in immature lungs.
ROC plot	A graphical depiction of diagnostic test accuracy data for all included studies.
Sensitivity	The ability of a diagnostic test to correctly identify women for whom delivery is imminent (within 48 hours or 7 days). Sensitivity = True Positive / (True Positive + False Negative)
Specificity	The ability of a diagnostic test to correctly identify women for whom delivery is not imminent (within 48 hours or 7 days). Specificity = True Negative / (True Negative + False Positive)
Single-gate study	Study design where participants' disease status is unknown and the index test result is evaluated against the reference standard to confirm the diagnosis.
Test failure	Rate of non-informative test results
Time to test	Time required to obtain test results
Tocolytic therapy	Drugs administered to delay the onset of established preterm delivery to allow time for in-utero transfers. Tocolytic therapy was previously used to allow a time to complete corticosteroid administration, however this is no longer recommended practice.

Plain English summary

Background

Infants may suffer from health problems if they are born early. If a mother has symptoms of labour before her baby is due, a test could be used to predict if the symptoms are real or a false alarm. A test would help the doctor decide whether the mother needs treatment or to move to a specialist hospital or could be sent home (if a false alarm).

Our report compares three tests (PartoSure, Actim Partus and Fetal Fibronectin) for how well they predict an early birth and how the costs and the ’s health compare between tests.

Methods

All the published literature reporting the accuracy of the three tests and their costs were found.

A cost-effectiveness model was written, which estimated the long-term health outcomes of the child based on the test results.

Results

Twenty studies were found which looked at how good the tests were at predicting an early birth within the next 7 days and six were found that looked at predicting birth within 48hours. The designs of the studies and the women taking part in the studies varied greatly. This meant that comparing the accuracy of the tests was very difficult and it would be unfair to decide which test was the best.

Our model suggested no firm conclusions for the cost-effectiveness of Fetal Fibronectin versus Actim Partus. PartoSure appear to be less costly than Actim Partus and equally good at predicting preterm birth but this is based on a study of very few patients. No studies allowed us to compare all three tests together.

Discussion and Conclusion

The accuracy of the results are uncertain, mainly because all the studies are very different. We are also aware of four related UK trials which are currently ongoing and plan to include large numbers of women.

Scientific summary

Background

Preterm (premature) birth, as defined by the World Health Organisation, refers to babies born alive before 37 weeks. Approximately 8% of births in England and Wales are premature. Preterm birth can result in serious short-term health issues in the infant, including difficulties with breathing (respiratory distress syndrome, RDS), feeding, increased risk of infections and bleeding within the brain (intraventricular haemorrhage, IVH). Moreover, long-term problems include an increased risk of cerebral palsy, cognitive and visual impairment, and respiratory illnesses.

Current NICE guidelines (2015) recommend that women presenting with symptoms of preterm labour who have intact membranes should undergo a clinical assessment. If the clinical assessment suggests that the woman is in suspected preterm labour and she is 29+6 weeks pregnant or less, treatment for preterm labour is recommended. If the clinical assessment suggests that the woman is in suspected preterm labour and she is 30+0 weeks pregnant or more, then the following tests should be conducted:

1. A transvaginal ultrasound measurement of cervical length (positive if <15mm)
2. If transvaginal ultrasound measurement of cervical length is unavailable or unacceptable, then a fetal fibronectin (fFN) test should be carried out (positive if concentration ≥50ng/ml)

Accurate diagnoses of preterm birth could prevent unnecessary (or ensure appropriate) admissions into hospital or transfers to specialist units.

Objectives

The purpose of this report is to assess the following three biomarker diagnostic tests for their test accuracy, clinical and cost-effectiveness.

- PartoSure (Parsagen Diagnostics Inc.) – a point of care dipstick test that detects placental alpha microglobulin-1 (PAMG-1) in vaginal secretions
- Actim Partus (Medix Biochemica, distributed by Alere) – a point of care dipstick test which detects phosphorylated insulin-like growth factor binding protein-1 (IGFBP-1) in cervical secretions
- Rapid Fetal Fibronectin 10Q Cassette Kit (Hologic, from now on known as quantitative fFN unless otherwise specified) used with a threshold ≠50ng/ml – a point of care quantitative test which detects the concentration of fFN in cervicovaginal fluid.

This assessment comprises of three systematic reviews of published literature corresponding to:

- Diagnostic test accuracy studies of the biomarker tests in symptomatic women with intact membranes;
- Clinical effectiveness (end-to-end) studies of the biomarker tests for symptomatic women with intact membranes;
- Economic evaluations of the biomarker tests for predicting preterm birth for symptomatic women with intact membranes;

In addition to these reviews, an independent economic evaluation is conducted.

Review of test accuracy

Methods

A systematic review was performed to assess the diagnostic test accuracy of PartoSure, Actim Partus and quantitative fFN. Studies were identified by searching seven bibliographic databases, trial registries, web-searching and additional supplementary search methods. Studies were selected if they met the following criteria:

- Population: Symptomatic women with intact amniotic membranes
- Index tests: PartoSure, Actim Partus, Rapid fFN 10Q Cassette Kit at thresholds ≠50ng/ml
- Reference Standards: Preterm delivery within 48hrs or within 7 days,
- Comparators: clinical assessment of symptoms alone, qualitative fFN, or quantitative fFN at a threshold of 50ng/ml
- Outcomes: Primarily sensitivity, specificity, positive predictive value and negative predictive value

Titles and abstracts were independently double-screened for inclusion and disagreements were resolved by discussion. Studies meeting inclusion at title and abstract stage were double screened as full texts.

The methodological quality of each included study was assessed using QUADAS-2, data were extracted, tabulated and narratively synthesised. Where the data allowed, summary receiver operating characteristic (ROC) plots were generated and meta-analyses were conducted.

Results

Twenty studies met the inclusion criteria, sixteen studies assessed Actim Partus, four assessed PartoSure and two assessed quantitative fFN.

Sufficient evidence was only available for pooling the test accuracy data for Actim Partus and PartoSure against the 7-day reference standard and Actim Partus against the 48hr reference standard. However, there was substantial methodological, clinical and statistical heterogeneity between studies raising considerable uncertainty about the most valid estimate of accuracy for each index test.

Studies offering the greatest certainty when comparing between tests were those that assessed two or more different tests within the same population. We identified two such studies: APOSTEL-1 (2016) and Hadzi-Lega (2017). From APOSTEL-1, depending on the threshold used quantitative fFN was more or less sensitive and specific than Actim Partus. From Hadzi-Lega, there was little difference between the sensitivity and specificity of PartoSure and Actim Partus. No studies assessed quantitative fFN and PartoSure within the same population.

When looking at all the studies identified for each of the tests and the ranges of results, the magnitude of the substantial heterogeneity between the studies is clearly apparent. Against the 7-day reference standard for Actim Partus (n=16 studies), the study with the best overall sensitivity and specificity results was Tripathi (2016, 94.7% 95%CI 89.9%, 97.7% and 92.4% 95%CI 88.9%, 95.1%) whilst Cooper (2012) reported the worst (sensitivity 33.3% 95%CI 4.3%, 77.7% and specificity 74.1% 95%CI 69.1%, 78.6%). For PartoSure (n=4 studies), the study with the best overall sensitivity and specificity results was Bolotskikh (2017, 100.0% 95%CI 73.5%, 100.0% and 95.4% 95%CI 88.6%, 98.7%) whilst Werlen (2015) reported the worst (sensitivity 0.0% 95%CI 0.0%, 97.5% and specificity 97.5% 95%CI 96.8%, 99.9%). The low sensitivity, from Werlen (2015), is attributable to only one woman testing (falsely) positive within the sample of 41. Fetal Fibronectin at a threshold of 10ng/ml (n=2 studies) had a sensitivity range of 93.8% (95%CI 82.8%, 98.7%) to 95.7% (95%CI 87.8%, 99.1%) and a specificity range of 32.2% (95%CI 27.7%, 37.0%) to 42.3% (95%CI 36.5%, 48.4%); at a threshold of 200ng/ml sensitivity ranged from 70.8% (95%CI 55.9%, 83.0%) to 71.0% (95%CI 58.8%, 81.3%) and specificity from 78.6% (95%CI 74.3%, 82.5%) to 83.6% (95%CI 78.8%, 87.8%) and at a threshold of 500ng/ml sensitivity ranged from 29.2% (95%CI 17.0%, 44.1%) to 42.0% (95%CI 30.2%, 54.5%) and specificity ranged from 94.3% (95%CI 91.6%, 96.4%) to 95.7% (95%CI 92.7%, 97.8%). Given the large ranges between studies assessing the same test and the significant overlapping of confidence intervals, it would be premature to attempt to deduce which test was superior against the 7 day reference standard.

We were only able to assess Actim Partus (n=6 study) and PartoSure (n=1 study) against the 48hr reference standard, since no studies were identified that assessed quantitative fFN. Similar to the 7-day results, accuracy results for Actim Partus varied substantially across studies. Given also that there was only one PartoSure study, it would also be premature to attempt to deduce which test was superior against the 48hr reference standard.

Review of clinical effectiveness (end-to-end) studies

Methods

The same literature search and screening methods were used as for the review of diagnostic test accuracy to identify randomised controlled or controlled studies of the tests (PartoSure, Actim Partus or fFN at thresholds ≠50ng/ml). Studies could compare the tests with each other or with fFN at a threshold of 50ng/ml, or clinical assessment of symptoms alone. Clinical outcomes were sought.

Results

No eligible studies were identified.

Review of economic evaluations

A systematic review was undertaken to identify previous economic evaluations of PartoSure, Actim Partus, and quantitative fFN. The methodology was identical to that used for the systematic review of test accuracy (described above). From 2252 records, 63 full texts were assessed for eligibility. Only one suitable (but unpublished) study was identified (Gibson 2014).

Gibson (2014) modelled the cost-effectiveness of a 'treat-all' strategy, relative to testing with qualitative fFN to determine treatment. Based on their findings we calculated the incremental cost-effectiveness ratio (ICER) of treating all suspected cases of preterm labour with antenatal corticosteroids is £20,942 per quality-adjusted life-year (QALY) gained.

Gibson (2014) also compared the use of four different quantitative fFN thresholds (10, 50, 200, and 500 ng/ml). Based on their results, we also calculated that testing at 200ng/ml dominates testing at lower thresholds, due to treatment and resource costs saved. However, the ICER of testing at 200ng/ml, relative to a higher threshold of 500ng/ml, was found to be £10,415 per QALY gained. Therefore, our calculations may support the authors' conclusions that using a 200ng/ml threshold for quantitative fFN was the optimal testing threshold. However, due to the low number of false negative cases in the study, there is a high level of uncertainty in their results.

To provide a more thorough examination of the evidence on modelling approaches, studies that modelled diagnostic interventions for suspected preterm labour were also examined. Six different model structures were identified, and all utilised a decision tree. The only cost-utility model identified was developed for the 2015 NICE guidelines for preterm labour. In addition to the decision tree structure, this model also extrapolated diagnostic results to obtain long term health outcomes of the child. The remaining studies were either cost-minimisation or cost-effectiveness analyses

Other major design aspects in which the six models differed were:

- Length of time horizon
- Assumptions surrounding adherence to treatment following a particular test result
- Type of treatment administered

Two studies conducted cost-minimisation analyses (i.e. did not consider effectiveness in terms of quality of life). The first was a Canadian study that found testing with fFN added total costs of approximately US$4,000,000, relative to no testing. The second was a UK study that compared clinical examination alone to clinical examination with a fFN test. This study found that using fFN saved the NHS £23.88 per patient, where the additional test costs were offset by the savings in hospital resource costs from treating fewer women.

Three studies provided cost-effectiveness analyses. The first (UK) compared testing with fFN to a 'treat-all' strategy. This model was unique in allowing for less than 100% admission following a positive test result. However, it did not consider outcomes for false positives, or compute results based on gestational age. The second study (US) found that treating all patients avoided US$433,000 in costs per case of RDS, and US$1,300,000 per neonatal life saved (1999 prices). It differed from other models by explicitly modelling preterm birth within 48 hours of testing. The third study (Netherlands) measured a variety of adverse outcomes as a composite measure, but only up until time of discharge (or death).

The 2015 NICE guidelines model presented a 'what-if' analysis of various testing strategies against a 'treat-all' approach. This involved varying the sensitivity and specificity of a hypothetical test to find the optimal values at which a test would be cost-effective, given a £20,000 per QALY threshold. The model was unique in measuring long term outcomes by gestational age. We comment in detail on NICE's model in this report and concluded that it provided the most suitable structure upon which to base our own model.

Independent economic assessment

We developed a new model that adopted the best methodological published practice including that of the 2015 NICE guidelines model. It models testing outcomes as a decision tree structure and projects long term health outcomes many years in to the future. Unlike the NICE model, which assumed all treatment involved tocolysis, our model considers treatment with antenatal corticosteroids only. Use of tocolysis is only assumed in case of hospital transfer. This is based both on recent evidence and current practice. Key features of the model include:

- Accounting for costs and lifetime QALY loss for an infant as a result of mortality, IVH, or RDS; as well as the QALY loss to the mother in a scenario analysis
- Differentiating costs and benefits by gestational age
- Distinguishing between hospital levels, and therefore accounting for the costs of a transfer from a lower to higher level hospital
- Accounting for the costs and benefits of antenatal corticosteroids for treatment of preterm labour, and the cost of tocolysis for transfers

The structure of the model is described briefly as follows. A woman with intact membranes, between 24 and 36 weeks' gestation, presenting with signs and symptoms of preterm labour, and for whom transvaginal ultrasound is not available or acceptable, is tested using one of fFN, Actim Partus, or PartoSure. Regardless of the result, this woman can either:

1. Give birth (preterm) within 7 days of the test
2. Give birth (with gestational age < 37 weeks) more than 7 days after testing
3. Give birth (with gestational age $\geq$ 37 weeks) more than 7 days after testing

If a woman tests positive, she is treated with steroids. If gestational age is below 28 weeks, and she presents at a level 1 or 2 hospital, she will also be given tocolysis and transferred to a level 3 (tertiary) hospital. In addition to the three tests, the model also considers a 'treat-all' strategy for comparison.

A review of health-related quality of life studies for preterm labour informed the selection of utilities for preterm survivors, IVH, RDS, and mothers. Due to a lack of suitable data in the literature, we use proxy utility values for IVH and RDS. Since only one study provided data for the quality of life of mothers who had previous adverse pregnancy outcomes, we do not include their utility as part of the base case. Overall, we improve on the utility data used in the NICE guidelines.

A review of cost studies informed the selection of relevant costs for inclusion in the model. Unlike the economic analysis that informed the NICE guidelines, our model accounts for the additional costs of saving a preterm neonatal life.

The results from our base case analysis (for a woman presenting at 30 weeks' gestation) are as follows. Using test accuracy data from Bruijn et al., we find that Actim Partus is £346 cheaper and 0.006 QALYs less effective than fFN at 50 ng/ml. This results in an ICER for Actim Partus of £56,030 cost saving per QALY lost vs fFN 50 ng/ml. Using test accuracy data from Hadzi-Lega (2017), we find that PartoSure is less costly than Actim Partus whilst being equally effective. No study provided data that allowed us to compare all three tests simultaneously. Indirectly comparing PartoSure to fFN 50 ng/ml (using APOSTEL-1 and Hadzi-Lega (2017)) yields an ICER of £81,922. Again, this represents both a cost saving and a QALY reduction and is highly uncertain given the indirect comparison source and the small size of the study by Hadzi-Lega (2017).

Discussion and Conclusion

There is a high degree of uncertainty surrounding the test accuracy results, primarily as a result of the substantial methodological, clinical and statistical heterogeneity between included studies. Nevertheless, our results suggest that the NICE Guideline recommendation that symptomatic women presenting at 30 weeks' gestation be admitted to hospital (i.e. the no test, treat all policy) may not be cost-effective. We are also aware of four ongoing UK trials, two of which are planning to enrol over 1000 participants (QUIDS and PETRA), whose results are likely to affect these conclusions.

Registration

PROSPERO: CRD42017072696

Word Count

2399 words.

1 Background and definition of the decision problem(s)

1.1 Condition(s) and aetiology(ies)

Preterm (premature) birth, as defined by the World Health Organisation (WHO), refers to babies born alive before 37 weeks and 0 days (37+0 weeks).[2]

Preterm birth can be serious for an infant in terms of both short and long-term health problems for the child and an increased risk of mortality. For example, short-term problems include difficulties with breathing (respiratory distress syndrome, RDS), feeding, an increased risk of infections and bleeding within the brain (intraventricular haemorrhages, IVH). Meanwhile, long-term problems include an increased risk of cerebral palsy, cognitive and visual impairment and respiratory illnesses.[3, 4]

1.1.1 Aetiology, pathology and prognosis

The WHO sub-categorises preterm birth based on gestational age as:[2]

- extremely preterm: <28 weeks gestational age (GA)
- very preterm: ≥28 weeks and <32 weeks GA
- moderate to late preterm: ≥32 weeks and <37 weeks GA

Iatrogenic preterm births are medically instigated deliveries, such as early labour induction or caesarean section.[5] These elective deliveries aim to reduce health risks to the mother or foetus due to complications such as hypertension, intrauterine growth restriction or pre-eclampsia.[5]

Spontaneous preterm labour is a multifactorial condition with various underlying pathologies including infection, breakdown of fetal-maternal tolerance, stress, decidual senescence and uterine distension (commonly associated with multifetal pregnancies).[6] Spontaneous preterm deliveries can be broadly categorised as either spontaneous labour with intact membranes or those following preterm premature rupture of membranes (PPROM).[5] Factors associated with an increased risk of preterm delivery include: stress, tobacco-use, drug-abuse, trauma, multifetal gestations, in vitro fertilisation, low BMI before pregnancy, extremes of maternal age, diabetes, high blood pressure, and infection.[7, 8] However, previous preterm delivery is the greatest risk factor for preterm birth.[9]

Symptoms of suspected preterm labour include: painful contractions or cramps, abdominal and low back pain, an increase or change in vaginal discharge.[10] Symptoms do not always result in progression to established labour and birth; they may occur but then settle allowing the pregnancy to continue towards term. It is understood that over 90% of women presenting

with symptoms of preterm labour do not go on to deliver in the next two weeks and of these 50% will continue with pregnancy until full term.[11, 12] It is important to determine whether preterm labour is the cause of the symptoms and assess the risk of preterm delivery, to allow appropriate management to begin as soon as possible.[13]

The focus population for this report are women presenting with signs and symptoms of spontaneous preterm labour with intact membranes.

1.1.2 Epidemiology

Data from the England and Wales 2016 birth cohort reports 54,143 live, preterm deliveries according to the WHO definition of preterm birth (<37 weeks gestational age), corresponding to 7.8% of total live births.[14] Of these deliveries 5.9% were categorised as extremely preterm (<28 weeks gestation), 10.4% were very preterm (gestational age ≥28 to <32 weeks) and 83.7% were moderate to late preterm (≥32 to <37 weeks gestation).[14]

2016 UK Birth cohort data collected by the Office for National Statistics show that the rate of preterm births vary between ethnic populations, with the highest proportion of preterm births affecting Black Caribbean and Indian populations (10.4% and 8.03% of pregnancies for these populations respectively) and the lowest rate of preterm births occurring in White Other (6.6%), while the rate of preterm delivery in the population where ethnicity was 'Not stated' was 8.3%.[14] In the UK preterm labour, particularly extreme preterm, disproportionately affects women from low socioeconomic backgrounds.[15, 16]

1.1.3 Incidence and/or prevalence

Improvements in perinatal healthcare services have resulted in vastly improved outcomes for babies born preterm, yet the prevalence of preterm birth continues to rise.[17, 18]

Preterm birth rates vary between countries, with higher prevalence and poorer outcomes in lower-income countries.[17] However, preterm birth is a global issue that also impacts developed countries.

1.1.4 Impact of health problem

Globally, preterm birth complications are directly responsible for 35% of all neonatal deaths and are the second leading cause of death in children under 5.[17, 19]

Morbidities associated with preterm birth are both acute and chronic and can affect all organ systems. Respiratory distress can progress to bronchopulmonary dysplasia,[20] and cerebral pathology e.g. intraventricular haemorrhages and ischaemia can lead to neurodevelopmental disorders including learning and behavioural difficulties.[21, 22] In addition

gastrointestinal disorders and immune-deficiencies, are also associated with preterm birth.[23, 24]

Although mortality and morbidity rates are higher for infants delivered at lower gestational age and lower birth weights, near-term premature infants remain at considerably higher risk of complications than their full-term counterparts.[22]

Preterm deliveries are a significant cost burden on the NHS. In addition to initial hospitalisation, re-hospitalisation and rehabilitation, other direct medical costs include medication, aids and devices such as wheelchairs, visits to physicians and home care.[25] Direct non-medical costs such as special education, adaptations to home or car, special meal requirements, higher insurance premiums and other disease-associated costs are an expensive burden on both families and the state.[25]

1.2 Current Guidelines

The NICE guideline on preterm labour and birth states that women reporting symptoms of preterm labour who have intact membranes should have a clinical assessment which includes:[26]

- clinical history taking
- observations of the woman, including: the length, strength and frequency of her contractions; any pain she is experiencing; pulse, blood pressure and temperature; and urinalysis
- observations of the unborn baby, including: asking about the baby's movements in the last 24 hours; palpation of the woman's abdomen to determine the fundal height, the baby's lie, presentation, position, engagement of the presenting part, and frequency and duration of contractions; and auscultation of the fetal heart rate for a minimum of 1 minute immediately after a contraction
- a speculum examination (followed by a digital vaginal examination if the extent of cervical dilatation cannot be assessed).

If the clinical assessment suggests that the woman is in suspected preterm labour and she is 29+6 weeks pregnant or less, treatment for preterm labour is recommended.[26]

If the clinical assessment suggests that the woman is in suspected preterm labour and she is 30+0 weeks pregnant or more then the following tests should be conducted:[26]

- Transvaginal ultrasound measurement of cervical length (as a diagnostic test to determine likelihood of birth within 48 hours).

- If cervical length is >15 mm, the woman is unlikely to be in preterm labour and could be discharged home with routine follow up in community and advised to return if symptoms reappear.
- If cervical length is ≤15 mm, the woman is diagnosed as being in preterm labour and should be offered treatment.

- If transvaginal ultrasound measurement of cervical length is indicated but is not available or not acceptable, then fetal fibronectin (fFN) testing as a diagnostic test may be used for women who are 30+0 weeks pregnant or more.
  - If fFN test result is negative (concentration <50 ng/ml), the woman is unlikely to be in preterm labour and could be discharged home with routine follow up in community and advised to return if symptoms reappear.
  - If fFN test result is positive (concentration ≥50 ng/ml), the woman is diagnosed as being in preterm labour and should be offered treatment

It is not recommended to use transvaginal ultrasound measurement of cervical length and fFN testing in combination to diagnose preterm labour.

1.3 Description of technologies under assessment

Accurate diagnoses of preterm birth using a biomarker test could prevent unnecessary or ensure appropriate admissions into hospital, transfers to specialist units and/or treatment.

1.3.1 Summary of the technologies

Following the NICE guidance, the technologies under assessment in this review would appear in the treatment pathway where the fFN test (at the threshold of 50ng/ml) is currently being used. A summary of information relating to the tests is given in Table 1.

1.3.1.1 PartoSure

PartoSure (Parsagen Diagnostics Inc.) is a CE marked qualitative lateral flow, immunochromatographic point-of-care test that detects placental alpha microglobulin-1 (PAMG-1) in vaginal secretions. PAMG-1 is protein produced by decidual cells lining the uterus and is secreted into amniotic fluid, its concentration in vaginal discharge is usually low and studies have shown the presence of PAMG-1 in vaginal discharge is predictive of imminent delivery.[27]

1.3.1.2 Actim Partus

Actim Partus (Medix Biochemica, distributed by Alere) is a CE marked qualitative immunochromatographic point-of-care test that detects phosphorylated IGFBP-1 (insulin-like

growth factor binding protein-1) in cervical secretions. Phosphorylated IGFBP-1 is made by cells lining the uterus and leaks into the cervix when delivery is imminent.[13]

1.3.1.3 Rapid Fetal Fibronectin 10Q Cassette Kit

The rapid fetal fibronectin 10Q Cassette is a CE marked point-of-care test for use in the PeriLynx System or the Rapid fFN 10Q System. This test quantifies the concentration of fFN present in cervicovaginal fluid. Fetal Fibronectin is a glycoprotein that connects membranes of the uterus and fetal membranes, which begins to degrade after 35th week of pregnancy or soon before preterm birth.

Table 1 Summary of Index tests

	Actim Partus[28, 29]	PartoSure (Parsagen Diagnostics Inc.)[30]	fFN[31]
Gestational Age	From 22 weeks	From 20+0 weeks to 36+6 weeks	From 22+0 weeks to 35+6 weeks
Contraindications	Ruptured membranes, vaginal bleeding (moderate or heavy), amniotic fluid.	Significant blood on the swab, within 6 hrs of vaginal disinfectant solutions or medicines. Inaccurate results may be likely with previous placenta previa or digital exam or in presence of meconium, anti-fungal creams, suppositories, lubricants, moisturisers, talcum powder or baby oil.	Advanced cervical dilatation (≥ 3cm), ruptured membranes, cervical cerclage, placental abruption, placenta previa (moderate) or vaginal bleeding (heavy). Inaccurate results may be likely with sexual intercourse, digital cervical exam or vaginal probe ultrasound and bacteria, bilirubin and semen. A negative test result is still valid if in the presence of semen.
Instructions	1. Take a 10-15 second swab from the cervical OS[a] during sterile speculum examination before any other investigations 2. Swirl the swab vigorously in the specimen extraction solution for 10-15 seconds 3. Place the dipstick in the specimen extraction solution until the liquid reaches the result area. 4. Remove dipstick from sample solution and lay horizontally for 5 minutes. **Negative results** (one blue line) should be confirmed at 5 minutes: highly unlikely that patient will deliver within the next 2 weeks **Positive results** (two blue lines) can be read as soon as it becomes visible (if before 5 minutes). Risk of a pre-term delivery is elevated.	1. Open the solvent vial and place in a vertical position by holding by the cap with all the liquid at the bottom. 2. Remove sterile flocked swab from packaging and with the patient lying on her back and by holding the middle of the swab shaft, insert into the vagina until the fingers contact the skin (5-7cm deep) for 30 seconds. 3. Place the swab in the solvent vial and rinse by rotating for 30 seconds. 4. Insert test strip into the solvent for 5 minutes. **Negative results** (one line) should be confirmed at 5 minutes. **Positive results** (two lines) can be read as soon as it becomes visible (if before 5 minutes).	1. Perform daily analyser quality control 2. During speculum examination, collect swab sample from the posterior fornix of the vagina and transfer to the transfer tube. 3. Mix sample in transport tube prior to removing swab expressing as much liquid as possible from the swab to tube 4. Mix the patient sample by removing the fFN 10Q Cassette from the foil pouch, enter necessary information into analyser and inserting the cassette into analyser. 5. When prompted pipette 200µL of patient sample into sample application well 6. Wait 10 minutes (7 minutes of incubation and 2-3 minutes of analysis) 7. fFN concentration will be displayed
Kit components	• One sterile polyester swab for specimen collection. • One tube of specimen extraction solution (0.5 ml). This phosphate-buffered solution contains bovine serum albumin (BSA), protease inhibitors and preservatives. • One dipstick in a sealed aluminium foil pouch with desiccant.	• One sterile flocked vaginal swab for specimen collection • One plastic vial with solvent solution. Solution contains 0.9% NaCl, 0.05% NaN_3 and 0.01% Triton X100 • One PartoSure test strip in foil pouch with desiccant.	• Rapid fFN 10Q Cassette Kit • PeriLynx Analyzer, Printer, User Manual, and PeriLynx QCette or Rapid fFN 10Q Analyzer, Printer, User Manual, and Rapid fFN 10Q QCette • Rapid fFN Control Kit • 200 µL pipette
Cost	£15 per test excl. VAT	£32 per test excl. VAT.	£35 per test excl. VAT per test.
Storage	The kit should be stored between 2 and 25°C	The kit should be stored in a dry place between 4 to 25°C	The kit should be stored at room temperature between 15° to 30°C. Transport specimens at 2° to2 5°C, or frozen. Specimens are stable for up to 8 hrs at room temperature. Specimens not tested within 8hrs of collection must be stored refrigerated at 2° to 8 ºC and assayed within 3 days of collection, or frozen and assayed within 3 months to avoid degradation of the analyte. Specimens arriving frozen should be subject to a single freeze-thaw cycle only.

Test range	The test has a limit of detection of 10 µg/l and a measuring range of 10 to 8,000 µg/l	The test has a limit of detection of 1 ng/ml and a measuring range of 1 to 40,000 ng/ml	The test has a detection range from 0 to 500ng/ml, concentrations greater than 500ng/ml will be displayed as >500 ng/ml
User personnel	The test is intended for professional use and results must be interpreted in the light of other clinical findings	The test is designed to be used in conjunction with clinical assessment and by healthcare professionals.	The test is intended to be used in conjunction with other clinical information

Note: a, Cervical OS, the opening of the uterine cervix.

1.3.2 Population

For the purpose of this report the population of interest are women with signs and symptoms of preterm labour with intact amniotic membranes, who are not in established labour and for whom a transvaginal ultrasound is not available or acceptable.

1.3.2.1 Identification of important sub-groups

The following women are at different risks of preterm delivery and consequently adverse neonatal outcomes. The clinical utility of the test may vary across these groups and the relative value of accurate identification of true positive and true negative cases is different from that in the overall population (NICE 2015).[26]

- Women with history of preterm delivery
- Women presenting with symptoms <28 weeks
- Women presenting with symptoms ≥28 and <32 weeks
- Women presenting with symptoms ≥32 weeks
- Women with multiple foetuses
- Women from lower socioeconomic groups (i.e. in most disadvantaged decile)

1.3.3 Current usage in the NHS

Current NICE guidelines are described in section 1.2.

Advising clinicians report that the guidelines are not always followed in typical clinical practice. Symptomatic women presenting irrespective of gestational age will usually have the fFN test administered. Additionally, clinicians advise transvaginal ultrasound is rarely used in routine practice either due to lack of trained staff available, experience or equipment availability.

1.3.4 Anticipated costs associated with the intervention

The cost of the Rapid fFN 10Q System is usually £35 per test, not including VAT. The cost per control is £40 and this is usually incurred twice per year for each site. Additional costs associated with equipment maintenance and test consumables are negligible. (Hologic 2017 Request for Information from NICE).

The cost per Actim Partus test is £15, not including VAT. No other costs are associated with this test. (Alere 2017 Request for information from NICE).

The cost per PartoSure test (PAMG-1) is £32, not including VAT. No other costs are associated with this test. (Parsagen 2017 Request for Information from NICE).

1.4 Comparators

The two comparators from the NICE Scope (for the clinical and cost effectiveness reviews) are fFN, used at a threshold of 50ng/ml and clinical assessment.

1.4.1 Fetal Fibronectin used with a threshold of 50 ng/ml

Point-of-care, qualitative fFN tests currently in use in the UK include QuikCheck fFN and Rapid fFN for the TLliq system.[32]

1.4.1.1 QuikCheck fFN

QuikCheck fFN (Hologic) is a CE marked, lateral flow immunoassay. The test kit includes a sterile applicator, test strip and tube containing extraction buffer. Additional materials required are a test tube rack and timer.[32]

Specimen is obtained from the posterior fornix using the applicator provided. The tip of the applicator is inserted into the extraction buffer and vigorously mixed for 10-15 seconds, applicator tip is pressed against the side of the tube to remove as much liquid as possible and discarded. The 'dip area' of the test strip is suspended in the extraction mixture for 10 minutes and then removed. Two lines indicates a positive result and high-risk of preterm delivery within 7-14 days; one line indicates a negative result and low-risk of delivery within 7-14 days; if no lines appear test result is invalid. The detection limit of the test is 50ng/ml.[32]

QuikCheck fFN test must be run within 15 minutes from sample collection. Sample should be obtained before digital examination is conducted as cervix disruption may affect test result. Presence of semen and gross vaginal bleeding may affect test result. The test is indicated for women presenting with threatened preterm labour and intact amniotic membranes.[32]

1.4.1.2 Rapid fFN for the TLIIQ System

Rapid fFN for the TLIIQ System (Hologic) is a CE marked immuno-chromatographic assay. Rapid fFN test kit includes cassettes and directional insert. Other materials required include 200µL pipette, Rapid fFN Control Kit (includes positive control, negative control and directional insert) and the TLIIQ System which contains analyser, printer and TLIIQ QCette.[32]

Cervicovaginal sample is obtained from the posterior fornix or the ectocervical region of the external cervical os using a swab. Swab is rolled against inside of the Specimen Transport Tube to express the liquid into the extraction buffer and swab is then discarded. TLiIQ Analyser is set to Internal Incubation Mode and the cassette-containing sample is inserted. 200µL of patient sample is dispensed into the sample application well of the Rapid fFN

Cassette. After 20 minutes the TLiIQ Analyser will display a result: positive, negative or invalid. The detection limit of the test is 50ng/ml.[32]

This test is indicated for use in routine prenatal visits between 22+0 weeks and 30+6 weeks' gestation in women, to assess the risk of delivery ≤ 7 or ≤ 14 days from testing. Disruption to the cervix i.e. through sexual intercourse, digital vaginal exam or vaginal probe ultrasound, may result in a false positive result. Douches, semen, white blood cells, red blood cells, bacteria and bilirubin may interfere with test result. However if the patient reports sexual intercourse within previous 24 hours, a negative fFN test result is still valid.[32]

1.4.2 Clinical assessment of symptoms alone

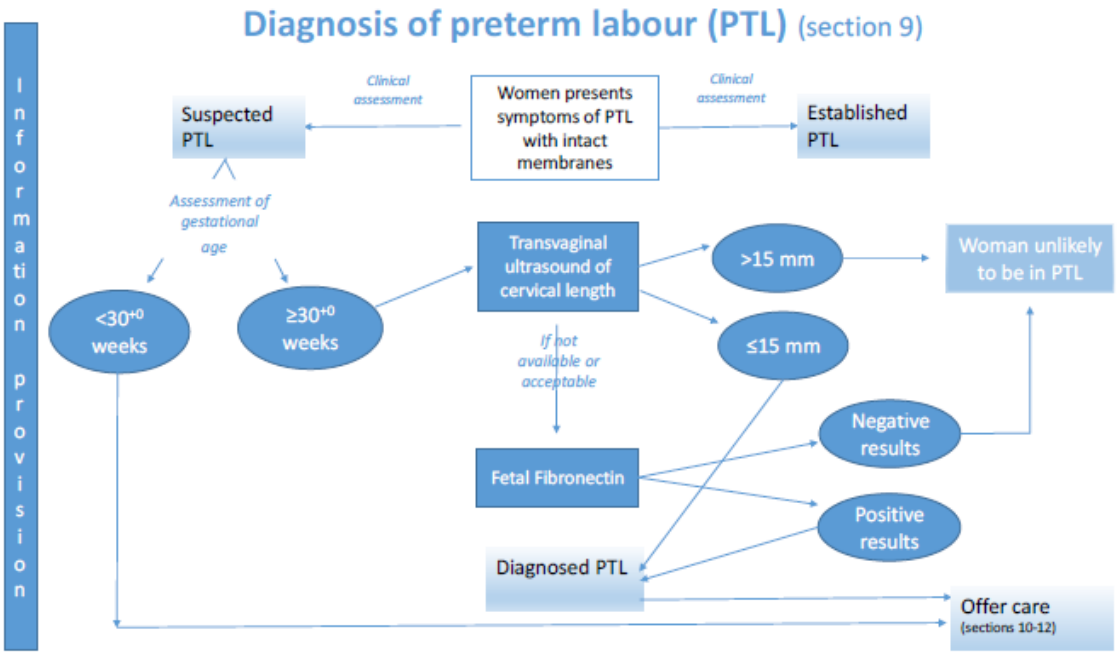
Clinical assessment consists of taking a clinical history, observations of the woman and unborn baby and a speculum exam. See section 1.2 for more details.

1.5 Care pathways

Clinical assessment and use of the biomarker tests aids clinicians in their decisions as to whether women presenting with signs and symptoms of preterm labour can be safely sent home or need to be admitted to hospital for treatment to delay birth and improve neonatal outcomes.[26] Typically, the results would be used in combination with clinical judgement, for example:

- If the test result is negative and the symptoms of preterm labour have settled, the woman would be discharged home with routine follow up in community and advised to return if symptoms reappear.
- If the test result is negative but symptoms of preterm labour continue, the woman would be admitted and monitored, and symptoms treated as appropriate and monitored. If symptoms were managed successfully, the woman would be discharged home.
- If the test result is positive, the woman would be admitted, and symptoms managed as appropriate and monitored.

Figure 1 Diagnosis of preterm labour from the 2015 NICE Guidance on Preterm Labour and Birth [26]



Once a woman has been diagnosed with threatened preterm labour, she will typically be offered tocolytic therapy, corticosteroids and magnesium sulphate.[26]

1.5.1 Tocolytic Therapy

Tocolytic therapies increase latency period for up to 48 hours, the aim of this therapy is to allow time for neonatal transfers and to complete the course of antenatal corticosteroids.[33] There are many classes of tocolytic drugs with different mechanisms of action.[33]

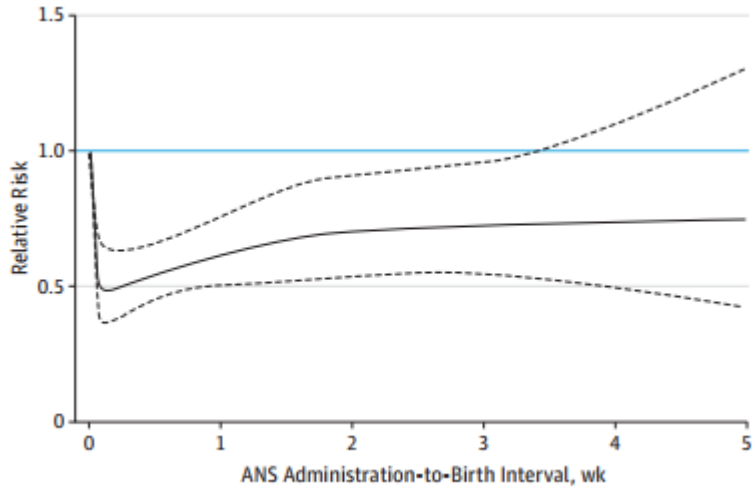
NICE guidelines recommend nifedipine for women between 24+0 and 33+6 weeks gestational age, in suspected or diagnosed labour with intact membranes. If nifedipine is contraindicated NICE recommends oxytocin receptor antagonists (e.g. Atosiban) for tocolytic therapy.[26]

Our clinical advisors suggest that tocolytic therapy is not commonly used in routine clinical practice. This may be due to recent evidence on the potential harms to the foetus and infant.[33, 34]

1.5.2 Antenatal Corticosteroids

Antenatal corticosteroids (e.g. dexamethasone or betamethasone) are prescribed in cases of threatened preterm labour to stimulate fetal lung development to reduce infant mortality and morbidity.[35] Following administration of steroids, there is a window within which the steroids appear to be most beneficial for the infant (Figure 2).[36] The primary documented negative effect of giving steroids is a negative reduction in birth weight of approximately 100g.[37]

Figure 2 Association between timing of antenatal corticosteroids and in-hospital mortality



Notes: The reference relative risk represents very preterm infants who were not exposed to ANS before birth. The curves for adjusted relative risks following ANS administration represent means and 95% CI bands.
Source: Norman et al. 2017[36]

NICE currently recommend antenatal corticosteroids for women between 26+0 and 33+6 weeks gestation in suspected, diagnosed or established preterm birth, PPROM or undergoing planned preterm delivery. Antenatal corticosteroids should be considered for extremely preterm (between 24+0 and 25+6 weeks gestation) and near-term (34+0 and 35+6 weeks gestation) women in suspected, diagnosed and established preterm labour, PPROM or undergoing iatrogenic deliveries.[26] However, evidence regarding the effectiveness of corticosteroid treatment at very low gestational age remains uncertain.[26] NICE recommends clinicians should discuss the benefits and risks associated with antenatal corticosteroids, with the patient and family. Repeat doses are contentious due to possible risks although this needs to be weighed against potential benefits of reduced RDS and serious adverse infant outcomes.[35, 38] As such, some cases of repeat dosing may be acceptable depending on: interval since last dose, gestational age and likelihood of delivery within 48 hours.[26]

1.5.3 Magnesium Sulphate

Magnesium sulphate is a neuroprotective agent that significantly reduces neurological morbidities such as cerebral palsy in preterm infants.[39]

NICE recommends Magnesium sulphate for women 24+0 to 29+6 weeks gestation, in established labour or with iatrogenic delivery planned within 24 hours. Magnesium sulphate should also be considered for women between 30+0 and 33+6 weeks pregnant.

A 4g IV bolus dose of magnesium sulphate should be administered over 15 to 20 minutes, followed by an IV infusion of 1g per hour for 24 hours or until delivery. Patients should be routinely monitored for signs of magnesium toxicity.

1.6 Outcomes

The accuracy of biomarker testing for predicting preterm labour has been evaluated against the reference standard of preterm delivery within 48hrs or 7 days. Clinically important outcomes relevant to test accuracy include:

- Sensitivity: the probability of correctly identifying someone who will deliver preterm

$$\text{Sensitivity} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} = \frac{TP}{TP + FN}$$

- Specificity: the probability of correctly identifying someone who will not deliver preterm

$$\text{Specificity} = \frac{\text{True negative}}{\text{False positive} + \text{True negative}} = \frac{TN}{FP + TN}$$

- Likelihood Ratio (LR) is the likelihood of a given test result in a patient who has a preterm delivery compared to the likelihood of that same result in a patient who does not deliver preterm
    - Likelihood ratio for positive test result (LR+). This is how much more often a positive test occurs in people who do deliver preterm compared to those who do not.

$$\text{Postive LR} = \frac{\text{P}(\text{Test} + \text{ve} \mid \text{preterm})}{\text{P}(\text{Test} + \text{ve} \mid \text{not preterm})} = \frac{\text{Sensitivity}}{1 - \text{Specificity}}$$

    - Likelihood ratio for negative test result (LR−). This is how much less likely a negative test result is in people with preterm delivery compared to those without preterm delivery.

$$\text{Negative LR} = \frac{\text{P (Test} - \text{ve} \mid \text{preterm)}}{\text{P (Test} - \text{ve} \mid \text{not preterm)}} = \frac{1 - \text{Sensitivity}}{\text{Specificity}}$$

- Positive predictive value (PPV). This is the probability of someone with a positive result actually having a preterm delivery.

$$\text{PPV} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} = \frac{TP}{TP + FP}$$

- Negative predictive value (NPV). This is the probability of someone with a negative test result actually not having a preterm delivery.

$$NPV = \frac{\text{True Negative}}{\text{True Negative} + \text{False Negative}} = \frac{TN}{TN + FN}$$

- Diagnostic yield (also known as test positivity rate or apparent prevalence). This is the number of positive test results divided by the number of samples.
- Concordance. The proportion of cases in which the result of the test agrees with the clinical outcome.
- Prevalence. The proportion of women actually having a preterm delivery.
- Test failure (non-informative test result) rate.
- Time (required) to (obtain a) test result.

2 Assessment of test accuracy

2.1 Methods for reviewing test accuracy

The diagnostic accuracies of PartoSure, Actim Partus, and Rapid fFN 10Q Cassette Kit (at thresholds other than 50ng/ml, from now on known as quantitative fFN unless otherwise specified), were assessed by:

i) Conducting a systematic review of the research evidence for these three index tests. This review was undertaken following the general principles published by the University of York Centre for Reviews and Dissemination (CRD).[40] The protocol was registered on PROSPERO (CRD42017072696)

ii) Providing a non-systematic overview of the test accuracy of fFN at 50ng/ml (i.e. the test recommended in current NICE guidance in situations where cervical length is unavailable or unacceptable),[26] and comparing the test accuracies of the index tests with this additional data. In order to ensure that the tests were compared in the same populations, data from the non-systematic overview were taken only from studies included in the systematic review. This did not, therefore, constitute a systematic review of the non-index tests (fFN at 50ng/ml). In order to check that the data obtained for this non-index test were representative of the values more broadly available, systematic reviews of the test accuracy of fFN at 50ng/ml were also sought.

The methods used to produce the systematic review and the overview presented in this section, with the results presented in Section 2.2 (page 46 onwards) and a summary in Section 2.4 (page 90 onwards).

2.1.1 Methods of the systematic review

The aim of this systematic review was to identify and summarise the diagnostic test accuracy data for PartoSure, Actim Partus, and quantitative fFN (at thresholds other than 50ng/ml), from test accuracy studies that provide data for one or more of these index tests.

2.1.1.1 Identification of studies

2.1.1.1.1 Study sources and searches

To identify studies, the following bibliographic databases were searched from inception until July 2017: MEDLINE (R), MEDLINE(R) Epub Ahead of Print, MEDLINE In-Process & Other Non-Indexed Citations, MEDLINE (R) Daily and Embase (all via Ovid); CINAHL (via EBSCOhost); BIOSIS and Web of Science (via Clarivate Analytics); and the Cochrane Library (Cochrane Database of Systematic Reviews, CENTRAL, DARE, HTA and NHS EED;

all via Wiley Interface). The search strategies were developed by a senior information specialist (CC), and comprised of terms designed to identify the index tests. Methodological filters for test accuracy studies were not used to limit the study designs retrieved as these have been shown to reduce sensitivity[41] and also because the search results were used to screen studies for the other two reviews described in this report (see chapters 3 and 4). Search results were limited to English language studies. The full search strategies for each database are reproduced in Appendix 1. The search results were exported to Endnote X8 (Thomson Reuters, NY, USA) and de-duplicated using automatic and manual checking.

Additional sources were searched as follows:

- Systematic reviews identified by the bibliographic database searches were screened for includable studies. For the purpose of this review, a systematic review was defined as one that had: a focused research question; explicit search criteria that are available to view; explicit inclusion/exclusion criteria; sufficient data on included and excluded studies to populate a PRISMA flow diagram; a critical appraisal of included studies, including consideration of internal and external validity of the research; and a synthesis of the included evidence (narrative or quantitative)
- Trial registries were searched via the Clinical Trials.Gov website (https://clinicaltrials.gov/ct2/home) and ISRCTN (https://www.isrctn.com/editAdvancedSearch) using terms designed to identify the index tests (Appendix 1)
- Google Advanced was used to conduct web-searching (September 2017), using terms designed to identify the index tests (Appendix 1). For each term searched on Google Advanced, the first 50 hits were screened
- Items included after full-text screening were forward citation chased and screened using Scopus (Elsevier)
- The reference lists of included studies were screened
- The industry submissions to NICE were cross-checked for additional studies

2.1.1.1.2 Study selection

Relevant studies were screened in two stages. First, titles and abstracts returned by the search strategy were examined independently by two reviewers (two of JVC, SD, MB, HC) and screened for possible inclusion, using pre-specified inclusion and exclusion criteria (see section 2.1.1.2). Disagreements were resolved by discussion within the review team. Full texts of studies included at the title and abstract screening stage were obtained, as were full texts of studies identified from systematic reviews, from trial registry searches, from forward

and backward citation chasing, from references provided by the companies, and from web searching. Two researchers (two of JVC, SD, MB, HC) independently examined full texts for inclusion or exclusion. Disagreements were again resolved by discussion within the review team.

2.1.1.2 Inclusion and exclusion criteria

2.1.1.2.1 Population

In line with the NICE scope,[13] studies were included if they recruited pregnant women with signs and symptoms of preterm labour who were not in established labour and who had intact amniotic membranes. Studies were eligible regardless of whether they were based on samples that were high- or low-risk for preterm labour. Studies were also eligible regardless of whether it was stipulated that the recruited population had access to transvaginal ultrasound. There were no specific inclusion criteria relating to the number of weeks gestation of the women recruited, the study was required however to define their population as preterm. The unit of assessment was individual women with a single result for each test.

Initially, studies were only included if all participants were expecting a singleton pregnancy. However, due to the lack of evidence, a protocol amendment was made to include studies where twin or multiple pregnancies were included but made up ≤20% of the total population recruited. We are not aware of any published evidence to suggest that multi-foetal pregnancies would alter the diagnostic test accuracy of any of the tests.

2.1.1.2.2 Index tests

In accordance with the NICE scope,[13] the index tests to be considered were:

- PartoSure (with or without a clinical assessment)
- Actim Partus (with or without a clinical assessment)
- Rapid fFN 10Q Cassette Kit (quantitative fFN), used with a threshold other than 50 ng/ml (with or without a clinical assessment) from now on known as quantitative fFN unless otherwise specified.

Studies were eligible for inclusion if one or more index test was assessed against a reference standard.

2.1.1.2.3 Reference standard

Studies using one or more of the following reference standards were eligible for inclusion:

- Preterm delivery within 48hrs, or within 7 days
- Clinical assessment of symptoms alone

- fFN at a threshold of 50ng/ml (qualitative or quantitative test)

In addition, studies that provided test accuracy data by comparing the results of one index test against another (i.e. by using one of the index tests as a reference standard) were also eligible for inclusion. It was, however, expected that most studies would use preterm delivery within 48hrs, or within 7 days, as the reference standard.

2.1.1.2.4 Outcomes

In accordance with the NICE scope,[13] the outcomes assessed for index tests were:

- Sensitivity: True Positive/(True Positive + False Negative)
- Specificity: True Negative/(False Positive + True Negative)
- Likelihood ratio for positive test result (LR+)
- Likelihood ratio for negative test result (LR−)
- Positive predictive value (PPV):True positive/(True Positive + False Positive)
- Negative predictive value (NPV): True Negative/(True Negative + False Negative)
- Diagnostic yield (also known as test positivity rate or apparent prevalence)
- Concordance
- Prevalence (or incidence) of preterm delivery within 7 days and/or within 48hrs
- Test failure (non-informative test result) rate
- Time to test result

2.1.1.2.5 Study design

Single-gate prospective or retrospective diagnostic studies with random or consecutively recruited participants were considered the optimal design for evaluating test accuracy of the index tests and were, therefore, eligible for inclusion. Ideally, studies assessing two or more index tests in the same population were sought, but studies assessing the accuracy of only one index test were also included. Studies assessing an index test and a test out of scope would be eligible for inclusion providing data reports specifically for all women receiving the index test. Two-gate diagnostic studies were also eligible for inclusion.

Studies where the index test was conducted within 7 days of the reference standard were included. In addition, a protocol amendment was made to include studies using frozen samples (i.e. use not in line with clinical practice), even when the test was analysed outside of the window stipulated in the manufacturers' guidelines due to a lack of evidence.

Studies were eligible for inclusion in the DTA review whether or not the index test results were used in the clinical management of patients.

We did not consider unpublished data without sufficient study methodology for quality appraisal.

2.1.1.3 Data extraction strategy

Data were extracted by one reviewer (SD) using a standardised data extraction form and checked by a second reviewer (JVC). Disagreements were resolved by discussion, with involvement of a third reviewer (HC) as necessary. Data were then transferred to standardised tables.

2.1.1.4 Critical appraisal strategy

The methodological quality of the studies was assessed by one reviewer (SD) and judgements were checked by a second reviewer (HC), according to criteria specified by Phase 3 of the QUADAS-2 tool (Appendix 3).[42] Any disagreements were resolved by discussion, with involvement of a third reviewer (JVC) as necessary.

2.1.1.5 Methods of data synthesis

For all included studies, sensitivity, specificity, positive and negative predictive value, positive and negative likelihood ratio, prevalence, concordance and diagnostic yield for delivery within 48 hours and 7 days were calculated from true positive, true negative, false positive and false negative values. Where the raw values were not provided, they were derived using back calculation from other suitable available data. Summary receiver operating characteristic (ROC) plots were generated to provide graphical depiction of the sensitivity and specificity data. These were produced for each test separately against the 48 hour and 7 day reference standards, and for quantitative fFN were also produced separately for each testing threshold.

Summary ROC plots were generated subject to a minimum of three studies per plot. In accordance with Stata requirements, the minimum number of studies for a diagnostic meta-analysis was four. Whenever this requirement was met, consideration was given to conducting meta-analysis. According to the Cochrane DTA handbook, heterogeneity cannot be assessed for diagnostic meta-analysis in the same way as for meta-analysis of interventions, and no quantitative summary statistic for heterogeneity can be derived.[43]

Meta-analysis against the 7-day delivery reference standard for Actim Partus was conducted using the metandi command for sensitivity and specificity. Meta-analysis against the 48 hour delivery reference standard for Actim Partus and against the 7-day reference standard for PartoSure was conducted using a mixed effects multilevel logistic regression to refine the

parameters used in metandi to improve model convergence in the presence of a low proportion of preterm births in the study by Werlen et al[44]. No meta-analysis was undertaken for quantitative fFN at any threshold or for the 48 hour reference standard for PartoSure.

Stata version 14.11 (StataCorp, College Station, TX) software was used for all statistical analysis. Graphs were made using Stata or Review Manager version 5.3 (Nordic Cochrane Centre, Copenhagen) software.

2.1.2 Methods of the overview

In addition to the systematic review described in section 2.1.1, an overview of studies that assess the diagnostic test accuracy of at least one of the index tests (PartoSure, Actim Partus, and quantitative fFN), in addition to a qualitative fFN test and/or quantitative fFN at 50ng/ml was provided since these are the tests currently recommended by NICE guidance.[26]

These data were extracted, where available, from the studies included in the systematic review of PartoSure, Actim Partus, and quantitative fFN (section 2.2). As such, studies included in the systematic review that also provided test accuracy data for a qualitative fFN test and/or quantitative fFN at 50ng/ml were included in this overview. Test accuracy data for qualitative fFN and/or quantitative fFN at 50ng/ml were extracted, tabulated and analysed following the same methods and principals described in section 2.1.1 above, although only sensitivity, specificity, PPV and NPV are summarised (and no meta-analyses were conducted). These test accuracy data are compared (in tables and text) with the test accuracy data for PartoSure, Actim Partus, and quantitative fFN, obtained from the same studies.

It should be noted that, because only 'comparative' DTA studies are summarised (i.e. studies providing data for both an index test and one or more of qualitative fFN and/or quantitative fFN at 50ng/ml), this summary does not systematically cover the full breadth of DTA evidence on these latter tests.[1] In order to ensure that the test accuracy data for qualitative fFN and/or quantitative fFN at 50ng/ml included here are largely representative of all data available for these tests, recent systematic reviews of DTA studies for these tests were sought and assessed.

In a similar manner, a further overview incorporating cervical length is given in Appendix 3.

[1] These studies are not comparative in the strictest sense, rather they evaluate more than one test within the same population (but do not directly compare the tests). This applies throughout this chapter.

2.2 Results of the systematic review

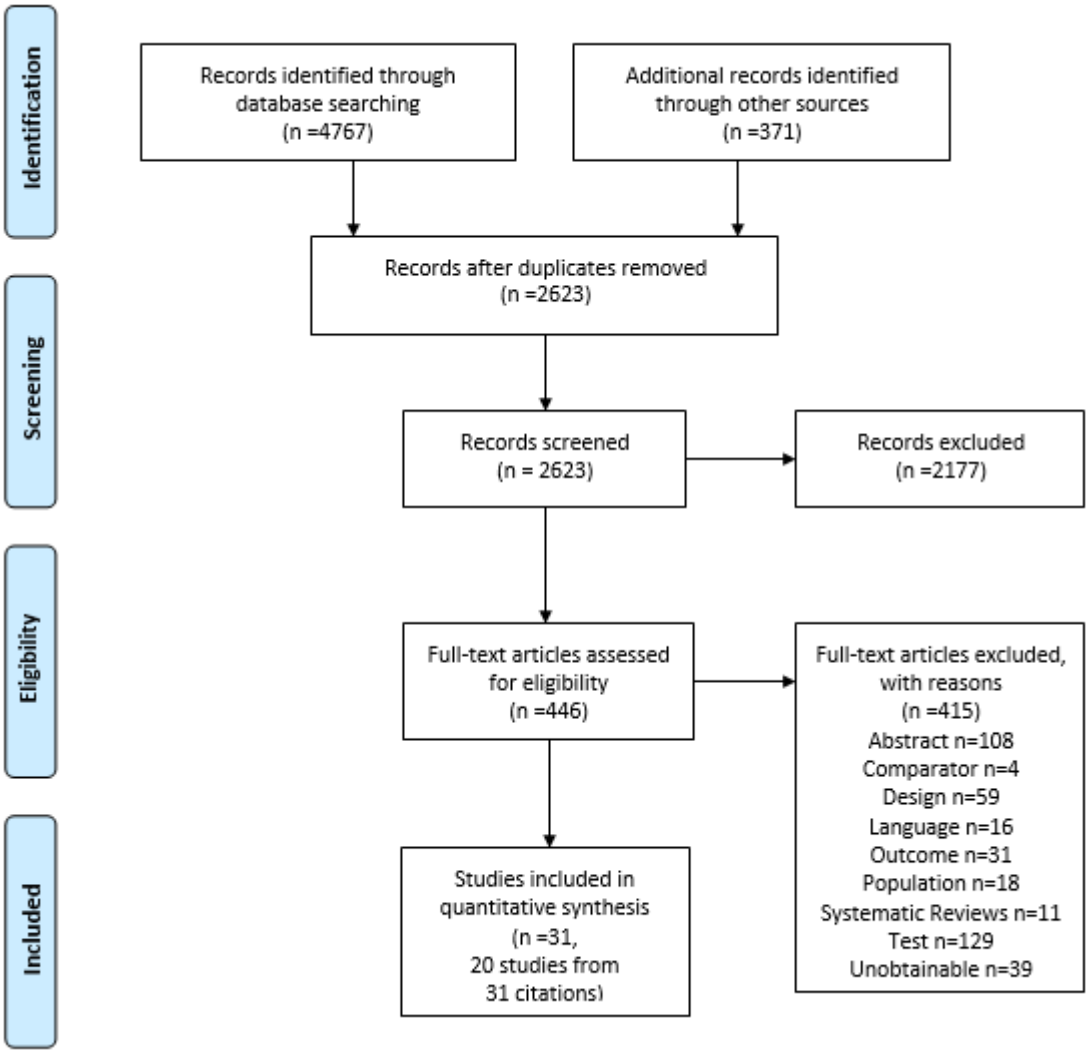
In this section, results of the systematic review of PartoSure, Actim Partus, and quantitative fFN are presented. Studies providing DTA data for one or more of these index tests are included (see 2.1.1.2 for specific details on inclusion and exclusion criteria).

2.2.1 Overview of the quantity and quality of research available

The searches retrieved a total of 2,619 unique titles and abstracts. A total of 2,177 articles were excluded, based on screening titles and abstracts. The remaining 442 articles were requested as full texts for more in-depth screening.

Of the 442 articles retrieved as full texts, 415 were excluded. The primary reasons for exclusion were: use of irrelevant test, typically qualitative fFN (n=129), the study design (n=59), outcomes did not match the review inclusion criteria (n=31), or the article was an abstract that had both insufficient information to be included in the review and was unconnected to any of the included studies (n=108). Abstracts were included if they were connected (by reporting data from the same study) to a full text included study. The bibliographic details of studies retrieved as full papers and subsequently excluded, along with the reasons for their exclusion are detailed in Appendix 2. Additional tables (Table 51, Table 52, Table 53) are provided in Appendix 2 listing all the citations provided by the industry to NICE along with whether the citation was included, and if not, the reason for exclusion. After screening relevant systematic reviews (n=11, see Table 49 in Appendix 2), and forward and backward citations of the included studies, no further new included studies were identified. Twenty studies from 31 citations met the review inclusion criteria. The process of study selection is shown in Figure 3.

Figure 3 Summary of the selection process



2.2.1.1 Ongoing trials

A search of trial registries and company submissions identified seven ongoing trials that may be relevant to this review of diagnostic test accuracy. These trials are summarised in Table 2.

Table 2 Ongoing trials

Study ID	Title	Sponsor	Status	Location	Estimated enrolment	Test(s)
NCT01987024	Advantage of Detection of phIGFBP-1 to Reduce Hospitalization Time for Stable Patients With a Risk of Preterm Labour.	Assistance Publique Hôpitaux De Marseille	Unknown	France	420	Actim Partus
NCT01868308	Screening To Obviate Preterm Birth (STOP)	University of Pennsylvania	Completed	United States	568	fFN
NCT02853656	Time to Delivery of Preterm Birth	Basildon and Thurrock University Hospitals NHS Foundation Trust	Currently recruiting	UK	242	Actim Partus and fFN
NCT02904070	Interest of Placental Alpha-microglobulin-1 Detection Test to Assess Risk of Premature Delivery in Reunion Island (PARTOSURE-OI)	Centre Hospitalier Universitaire de la Réunion	Currently recruiting	Réunion Island, France	300	PartoSure, Actim Partus and fFN
ISRCTN41598423	Can a test of preterm labour (quantitative fetal fibronectin) help diagnosis and clinical decision making? (QUIDS and QUIDS-2)	Health Technology Assessment Programme (UK)	Currently recruiting	UK	2100	PartoSure, Actim Partus and fFN
IRAS ID 111142	Threatened preterm labour: a prospective cohort study of a clinical risk assessment tool and a qualitative exploration of women's experiences of risk assessment and management. (PETRA)	King's College London	Currently recruiting	UK	1181	fFN
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Four of the ongoing trials are based in the UK, two of which are planning to enrol over 1000 participants (ISRCTN41598423 N=2100; IRAS ID 111142 N=1181). However, it was not possible to include data from these ongoing trials in this review of test accuracy.

2.2.2 Description of the included studies

Characteristics of the included studies are summarised in Table 3. Two studies, APOSTEL-1 (2016)[45, 46] and Hadzi-Lega (2017)[47] assessed the diagnostic test accuracy of two different

index tests in the same population; APOSTEL-1 assessed both Actim Partus and quantitative fFN while Hadzi-Lega (2017) assessed Actim Partus and PartoSure. APOSTEL-1 was one of the larger studies (n=350), conducted in ten centres around the Netherlands, while Hadzi-Lega (2017) was a smaller study (n=57) from one centre in Macedonia.[45-47]

A further 14 studies (Abo El-Ezz 2014, Altinkaya 2009, Azlin 2010, Brik 2010, Cooper 2012, Danti 2011, Eroglu 2007, Goyal 2016, Lembet 2002, Riboni 2011, Tanir 2009, Ting 2007, Tripathi 2016, and Vishwekar 2017) assessed the diagnostic test accuracy for only Actim Partus.[1, 48-60] Three studies (Bolotskikh 2017, Nikolova 2015 and Werlen 2015) assessed PartoSure only[44, 61-63] and one (EUIFS, 2016) assessed quantitative fFN only.[64]

For Actim Partus, study size ranged from n=30 in Vishwekar (2017) to n=468 in Tripathi (2016) and covered the following countries: Kuwait, Turkey, Malaysia, Spain, Canada, Italy, India, and Singapore.[1, 48-60] The three studies assessing PartoSure were conducted in Russia, Macedonia and France and the study size ranged from n=41 in Werlen (2015) to n=203 in Nikolova (2015).[44, 61-63] Finally, the EUIFS study, assessed quantitative fFN only, and recruited n=455 participants from 10 centres across Austria, Belgium, Germany, Netherlands and Switzerland.[64]

Table 3 Study Characteristics

Study	Other tests used	N Included, (Recruited)	Country (number of centres)	Definition of Pre Term Labour Symptoms	Weeks gestation	Dilation threshold for exclusion	Other exclusion criteria
Study assessing Actim Partus and quantitative fFN							
APOSTEL-1 [45, 46]	Cervical Length	350 (714)	Netherlands (10)	Uterine contractions (>3/30min), abdominal pain, back pain, vaginal bleeding	24-34	>3 cm	Contraindications for tocolysis, Iatrogenic deliveries, Tocolytic treatment prior to testing
Study assessing Actim Partus and PartoSure							
Hadzi-Lega, 2017 [47]	Cervical Length	57 (72)	Macedonia (1)	Uterine contractions, abdominal pain	22-34+6	>3 cm	Antepartum haemorrhage, Cervical cerclage, Multiple gestations
Actim Partus							
Abo El-Ezz, 2014 [48]	NA	57 (80)	Kuwait (2)	Uterine contractions (≥8/hr), back pain, pelvic pressure, vaginal discharge, 50% effacement	24-34	>3 cm	Cervical cerclage, Chorioamnionitis, Fetal abnormalities, Intrauterine growth restriction, Multiple gestations, Placenta praevia, Prior cervical examination, Sexual intercourse previous 24h, Uterine anomalies, Vaginal bleeding
Altinkaya, 2009 [49]	NA	105 (NR)	Turkey (1)	Uterine contractions	24-34	≥2cm	Fetal abnormalities, History of preterm delivery, Intrauterine growth restriction, Multiple gestations, Preeclampsia, Smokers, Uterine anomalies, Vaginal bleeding
Azlin, 2010 [50]	Cervical length	51 (51)	Malaysia (NR)	Uterine contractions	24-36	≥3cm	Abruptio placenta, Cervical cerclage, Cervical incompetence, Multiple gestations, Placenta praevia
Brik, 2010 [51]	NA	276 (325)	Spain (1)	Uterine contractions, abdominal pain, back pain, leaking of fluid, other	24-34	>3cm	Abruptio placenta, Cervical cerclage, Fetal abnormalities, Fetal distress, Vaginal bleeding, (active labour)
Cooper, 2012 [1]	Qualitative fFN (unclear which test)	349 (366)	Canada (2)	Symptoms judged by physician to be indicative of labour	24+0 – 34+6	NR	Antepartum Haemorrhage, Chorioamnionitis, (active labour)
Danti, 2011 [52]	Cervical length	60 (102)	Italy (1)	Uterine contractions (≥4/20mins)	24+0 – 32+6	>3cm	Abruptio placenta, Cervical cerclage, Fetal abnormalities, Intrauterine growth restriction, Multiple gestations, Placenta praevia, Preeclampsia, Uterine anomalies, Vaginal bleeding
Eroglu, 2007 [53]	QuikCheck fFN Cervical length	51 (51)	Turkey (1)	Uterine contractions (>10/hr)	24 - 35	≥3cm	Abruptio placenta, Fetal abnormalities, Intrauterine growth restriction, Multiple gestations, Placenta praevia, Preeclampsia, Sexual intercourse previous 24h, Uterine anomalies, Vaginal bleeding
Goyal, 2016 [54]	Cervical length	60 (95)	India (1)	Uterine contractions (>4/20mins), abdominal pain	24 - 36	NR	Fetal abnormalities, Fetal growth restrictions, Preeclampsia, Multiple gestations, Vaginal bleeding,

Lembet, 2002 [55]	NA	36 (36)	Turkey (1)	Uterine contractions (>10/hr)	20 – 36	n/a	Fetal abnormality, Intrauterine growth restriction, Preeclampsia, Multiple gestations, Uterine anomalies, Vaginal bleeding
Riboni, 2011 [56]	fFN by ELISA	210	Italy (2)	Uterine contractions (>10/hr)	24 – 34	>2cm	Fetal abnormalities, Multiple gestations, Placenta praevia, Prior cervical examination, Sexual intercourse previous 24h, Uterine anomalies, Vaginal bleeding
Tanir, 2009 [57]	NA	68 (121)	Turkey (1)	Uterine contractions (>4/20mins), changes in cervix, back pain, increased discharge	24-37	≥3cm	Asthma, Cervical cerclage, Diabetes mellitus, Digital examination previous 24h, Hyperthyroidism, Multiple gestations, Preeclampsia, Sexual intercourse previous 24h, Tocolytic treatment prior to testing, Vaginal bleeding, Vaginal douche previous 24h
Ting, 2007 [58]	Qualitative fFN (unclear which test)	94 (108)	Singapore (1)	NR	24-34	≥3cm	Cervical cerclage, Chorioamnionitis, Fetal asphyxia, Fetal abnormalities, Intrauterine growth restrictions, Multiple gestations, Placenta praevia, Preeclampsia
Tripathi, 2016 [59]	QuikCheck fFN	468 (550)	India (1)	Uterine contractions (>1/10mins), labour pains	28+1-36+6	>3cm	Blood-mixed cervical secretions, Diarrhoea, Prepartum haemorrhage, Previous preterm delivery, Sexual intercourse previous 24h, Urinary tract infection, Vaginal leakage
Vishwekar 2017[60]	NA	30 (NR)	India (1)	Uterine contracts, vaginal discharge	28-37	NR	Blood-mixed cervical secretions, Fetal distress, Hypertension, Intrauterine growth restrictions, (active labour)
PartoSure							
Bolotskikh, 2017 [61]	Cervical length	99 (100)	Russia (1)	Uterine contractions, abdominal pain, back pain, pelvic pressure, menstrual-like cramping, diarrhoea	22+0 - 36+6	>3cm	Maternal age less than 18yrs, Multiple gestations, Prior cervical examination, Placenta praevia, Symptoms unrelated to threatened preterm delivery e.g. trauma, Tocolytic treatment prior to testing, Vaginal bleeding
Nikolova, 2015 [62, 63]	Cervical length, QuikCheck fFN	203 (219)	Macedonia and Russia (2)	Uterine contractions, abdominal pain, pelvic pressure	20+0 – 36+6	>3cm	Cervical cerclage, placenta praevia, Maternal age less than 18 years, Multiple gestations,
Werlen, 2015 [44]	NA	41 (42)	France (1)	Uterine contractions, cervical changes	24-34	>3cm	Blood-mixed cervical secretions, Multiple gestations, Vaginal infection
Quantitative fFN							
EUIFS [64]	Cervical length	455 (484)	Netherlands Switzerland Belgium Germany Austria (10)	Uterine contractions (>3/30mins), abdominal pain, back pain, vaginal bleeding	24-34	>3cm	Contraindications for tocolysis, Fetal distress, Iatrogenic deliveries. Tocolytic treatment prior to testing, Triplet or higher gestations,

Key: fFN, fetal fibronectin; NA, not applicable; NR, not reported

2.2.2.1 Key differences between studies

It was notable that the prevalence rates of preterm birth differed greatly between studies (Table 6 and Table 7). In addition, there were differences between studies in the mode of delivery for included women. The participant inclusion/exclusion criteria also differed between studies. For example, although all studies included women presenting with symptoms of preterm labour with intact membranes, the definition of preterm (i.e. the number of weeks gestation) differed between studies. The inclusion/exclusion criteria of the studies also differed with regards to the presenting symptoms of the women, the proportion of women with singleton gestations, the risk status of included women, dilation thresholds applied, and other specific exclusion criteria.

2.2.2.1.1 Differences between studies in prevalence of preterm birth

There was clear variation between studies with regards to the prevalence of the reference standard (i.e. the prevalence of preterm birth within 7 days and within 48hrs). Across all 20 studies prevalence of preterm birth within 7 days ranged from 1.7% (95% CI 0.6, 3.7) in Cooper (2012) to 73.3% (95% CI 60.3, 83.9) in Goyal (2016).[1, 54] Both of these studies assessed the test accuracy of Actim Partus. The studies assessing Actim Partus, therefore, had a larger range of prevalence (of preterm birth within 7 days) than the studies assessing PartoSure and qualitative fFN. However, in the studies assessing PartoSure, prevalence of preterm birth with 7 days still ranged widely (from 2.4%, 95% CI 0.1, 12.9 in Werlen, 2015 to 17.2%, 95% CI 12.3, 23.2 in Nikolova, 2015). In the studies assessing quantitative fFN, prevalence of preterm birth within 7 days was slightly higher in APOSTEL-1 than in EUIFS (19.7%, 95% CI 15.7, 24.3 vs 10.5%, 95% CI 7.9, 13.7 respectively).

Seven studies provided DTA data for the index tests against preterm birth within 48hrs.[44, 51, 54, 55, 58-60] Across these seven studies, the prevalence of preterm birth within 48hrs ranged from 2.4% (95% CI 0.1, 12.9) in Werlen (2015) to 58.3% (95% CI 44.9, 70.9) in Goyal (2016).[44, 54] The study by Werlen (2015) was the only study assessing PartoSure against preterm birth within 48hrs, with the other six studies (Brik 2010, Goyal 2016, Lembet 2002, Ting 2007, Tripathi 2016, Vishwekar PS 2017) assessing Actim Partus.[44, 51, 54, 55, 58-60] The lowest prevalence of preterm birth within 48hrs within these six Actim Partus studies was 5.3% (95% CI 1.7, 12.0) in Ting (2009).[58]

These differences in prevalence displayed between studies are likely due to differences in the populations recruited into the studies (e.g. differences in gestational age, and in presenting symptoms of preterm labour, see section 2.2.2.1.3) and will also likely impact upon the DTA data presented in section 2.2.6 and the generalisability of these data to the NHS in England.

2.2.2.1.2 Differences between studies in mode of delivery

It is important to know whether women who had non-spontaneous deliveries within the time-frame of the reference standard were included or excluded from the test accuracy data; if iatrogenic delivery occurs within this timeframe it remains unclear whether a spontaneous delivery may have occurred, and thus makes it impossible to accurately assess the reference standard in these women. Nine of the included studies (Abo El-Ezz 2014, Altinkaya 2009, Danti 2011, Eroglu 2007, Goyal 2016, Riboni 2011, Ting 2007, Tripathi 2016 and Werlen 2015) did not report the mode of delivery (i.e. whether birth was spontaneous, or whether there were any planned caesarean sections or inductions).[44, 48, 49, 52-54, 56, 58, 59]

The other eleven studies (APOSTEL-1, Hadzi-Lega 2017, Azlin 2010, Brik 2010, Cooper 2012, Lembet 2002, Tanir 2009, Vishwekar 2017, Bolotskikh 2017, Nikolova 2015 and EUIFS) provided some data regarding mode of delivery (see Table 4).[1, 45-47, 50, 51, 55, 57, 60-64] Four of these studies (APOSTEL-1, Hadzi-Lega 2017, EUIFS, Bolotskikh 2017) reported that women who had a non-spontaneous delivery within the time-frame of the reference standard were excluded from the test accuracy data.[45-47, 61, 64] Women who had a non-spontaneous delivery outside of the time-frame of the reference standard should not be excluded as these women would be considered to be reference standard negatives in any case (i.e. they did not deliver within 48hrs or within 7 days). In a further three studies (Vishwekar 2017, Brik 2010, Nikolova 2015) iatrogenic delivery was mentioned as a reason for exclusion, but it is unclear how many of these deliveries occurred within the timeframe of the reference standard, and in another study (Lembet 2002) the number of iatrogenic deliveries could not be ascertained.[51, 55, 60, 62, 63]

In three studies (Azlin 2010, Cooper 2012, and Tanir 2009) the numbers of spontaneous/iatrogenic deliveries were reported, but no exclusion of data from non-spontaneous deliveries was made.[1, 50, 57, 120] In Azlin 2010, 70.6% of women delivered spontaneously, 7.8% underwent emergency caesarean section and for a further 7.8% there were no data available on mode of delivery. Although 13.7% delivered by a planned caesarean section, it is unclear whether these occurred within the timeframe of the reference standard, and these women were not excluded from the test accuracy data.[50] In Cooper (2012), 52.1% delivered spontaneously, 14.9% had an operative delivery, and 33.0% had a caesarean section, although again it was unclear how many were planned and how many occurred within the timeframe of the reference standard.[1] Finally, Tanir (2009) report the proportion of women who delivered either by caesarean or from vaginal delivery, however these data do not appear to be correct (see Table 4).[57]

2.2.2.1.3 Differences between studies in inclusion/exclusion criteria

There are several ways in which the participant inclusion/exclusion criteria differed between studies (see Table 3). It is likely that many of these differences impacted upon prevalence of preterm labour and thus test accuracy data. However, insufficient data were provided to fully assess these relationships.

Gestational age

Of the 20 included studies, 14 recruited women from 24 weeks gestation (Riboni 2011, APOSTEL-1, Altinkaya 2009, Brik 2010, Ting 2007, Werlen 2015, EUIFS, Abo El-Ezz 2014, Tanir 2009, Goyal 2016, Eroglu 2007, Danti 2011, Cooper 2012, Azlin 2010; see Table 3). In these 14 studies the upper gestational age varied widely: eight recruited women with gestations up until 34 weeks (Riboni 2011, APOSTEL-1, Altinkaya 2009, Brik 2010, Ting 2007, Werlen 2015, EUIFS, and Abo El-Ezz 2014), with the upper gestation age ranging from 32 +6 weeks (Danti, 2011) to 37 weeks (Tanir 2009) in the remaining six studies.[1, 44-46, 48-54, 56-58, 64]

Of the remaining six studies, four recruited women from an earlier gestational age: two studies (Lembet 2002 and Nikolova 2015) recruited women from as early as 20 weeks gestation and two studies (Hadzi-Lega 2017 and Bolotskikh 2017) included women from 22 weeks. Two of these studies recruited women up until 36+6 weeks (Nikolova 2015, Bolotskikh 2017), one until 36 weeks (Lembet 2002) and the other until 34+6 weeks gestation (Hadzi-Lega 2017).[47, 55, 61-63] The other two studies (Tripathi, 2016 and Vishwekar, 2017) recruited women from a later gestational age (i.e. from 28 weeks gestation) with both recruiting up until 37 weeks (36+6 for Tripathi 2016).[59, 60]

None of the studies presented test accuracy data between different gestational cut-offs. It was not possible, therefore, to make any within-study assessment, for any of the index tests, as to whether test accuracy differed based on gestational age.[1, 44-64]

Presenting symptoms of preterm labour

Other than stating that women had to be symptomatic, all studies except for Ting (2007), provided some further details about the presenting symptoms of preterm labour. However, one further study (Cooper 2012) only added that physicians determined "*symptoms indicative of labour were to be determined by a physician.*[1, 58]

All other studies reported uterine contractions as a necessary indicator of preterm labour.[44-57, 59-64] Ten studies additionally described the rate of uterine contractions necessary for inclusion: six contractions per hour in Tripathi (2016), EUIFS (2016) and APOSTEL-1 (2016),[45, 46, 59, 64] eight contractions per hour in Abo El-Ezz (2014) [48], ten contractions an hour

in Eroglu (2007), Riboni (2011) and Lembet (2002),[53, 55, 56] and 12 contractions per hour in Danti (2011), Goyal (2016) and Tanir (2009).[52, 54, 57]

Other commonly reported symptoms included in definitions of preterm labour were: abdominal pain in seven studies (Nikolova 2015, Bolotskikh 2017, Hadzi-Lega 2017, Goyal 2016, Brik 2010, EUIFS and APOSTEL-1),[45-47, 51, 54, 61-64] back pain in six studies (Abo El-Ezz 2014, Bolotskikh 2017, Brik 2010, Tanir 2009, EUIFS 2016 and APOSTEL-1 2016),[45, 46, 48, 51, 57, 61, 64] pelvic pressure in three studies (Nikolova 2015, Bolotskikh 2017 and Abo El-Ezz 2014),[48, 61-63] vaginal bleeding in the APOSTEL-1 and EUIFS studies,[45, 46, 64] and vaginal discharge in three studies (Abo El-Ezz 2014, Tanir 2009 and Vishwekar 2017).[48, 57, 60]

Singleton/multiple pregnancies

The majority of included studies were based on samples that only included women with singleton pregnancies. However, based on the protocol amendment, four studies were included that recruited women with multiple gestation pregnancies. One of these studies assessed two index tests in the same population APOSTEL-1 (quantitative fFN and Actim Partus) and included 20% multiple pregnancies.[45, 46] Two of these studies assessed only Actim Partus, with one (Cooper 2012) including 6% and the other (Vishwekar 2017) 7% multiple pregnancies.[1, 60] The final study that included multiple pregnancies (EUIFS) only assessed quantitative fFN, and they made up 15% of the population.[64] We are not aware of any published evidence to suggest that multi-foetal pregnancies would alter the diagnostic test accuracy of either quantitative fFN or Actim Partus.

Risk status of participants

Only one of the included studies (Bolotskikh 2017) clearly reported the risk status of included women (i.e. whether or not the women were high or low risk for preterm labour prior to the onset of symptoms).[61] In this study, the population was described as high risk because 15% had previously experienced preterm labour, 43% had mild preeclampsia and 51% were previously hospitalised during the pregnancy.[61] However, although none of the other studies explicitly stated that their populations were high risk for preterm labour, eight additional studies (APOSTEL-1, Brik 2010, Cooper 2012, Eroglu 2007, Goyal 2010, Lembet 2002, Vishwekar 2017, EUIFS) also recruited some women who had previously experienced preterm delivery (see Table 4).[1, 45, 46, 51, 53-55, 60, 64] In addition, two studies (APOSTEL-1 and Danti 2011) restrict their population by performing tests in women presenting with a cervical length <30mm. However, this high-risk status is associated with symptoms at presentation rather than the women being high-risk prior to the onset of symptoms.[45, 46, 52]

Recruiting high-risk women would be expected to impact upon the prevalence of preterm birth. However, because almost all studies did not clearly report risk status it is not possible

to properly assess whether or how this impacted upon prevalence rates (and, therefore, test accuracy data).

Dilation threshold and cervical length

All studies except Cooper (2012), Goyal (2016), Lembet (2002) and Vishwekar (2017) included a dilation threshold for exclusion, typically the threshold was >3cm or ≥ 3cm.[1, 54, 55, 60] However, Riboni (2011) had a dilation threshold of >2cm and Altinkaya (2009) had a threshold of ≥2cm.[49, 56]

Studies were not excluded on the basis of access to cervical length measurement (lack of access was unlikely to be reported in studies). Indeed, studies that did not report the use of cervical length measurement did not explicitly cite lack of access or discuss the suitability of cervical length measurement for the included population. Cervical length measurement was conducted in nine of the included studies (APOSTEL-1, Hadzi-Lega 2017, Azlin 2010, Danti 2011, Eroglu 2007, Goyal 2016, Bolotskikh 2017, Nikolova 2015, EUIFS).[45-47, 50, 52-54, 61-64] In seven of these studies (Hadzi-Lega 2017, Azlin 2010, Eroglu 2007, Goyal 2016, Bolotskikh 2017, Nikolova 2015, EUIFS), no selection of women occurred according to cervical length measurement, and therefore, it is not expected that the women in these studies would substantively differ from women who would not have access to cervical length measurement in clinical practice.[47, 50, 53, 54, 61-64] However, in the other two studies (APOSTEL-1, Danti 2011), all women included in final analyses of index test data had a transvaginal cervical length measurement ≤30mm, which would likely increase the prevalence of preterm birth in these studies.[45, 46, 52] Both of these studies were assessing Actim Partus, with APOSTEL-1 additionally assessing quantitative fFN.[45, 46, 52]

Other exclusion criteria

Other criteria for exclusion differed substantially between studies (see Table 3 for specific details). Across studies, exclusion criteria included; abruptio placenta, antepartum haemorrhage, contraindications for tocolysis, cervical cerclage, cervical incompetence, chorioamnionitis, diabetes mellitus, diarrhoea, digital examination in the previous 24h, fetal asphyxia, fetal abnormalities, fetal distress, history of preterm delivery, hypertension, hyperthyroidism, iatrogenic deliveries, intrauterine growth restriction, maternal age less than 18yrs, multiple gestations, placenta praevia, preeclampsia, prior cervical examination, sexual intercourse in the previous 24h, smokers, symptoms unrelated to threatened preterm delivery (e.g. trauma), tocolytic treatment prior to testing, urinary tract infection, uterine anomalies, vaginal bleeding, vaginal douche in the previous 24h, vaginal infection and vaginal leakage.[1, 44-64] All exclusion criteria were reasonable in the context of the index tests under consideration (see section 2.2.5.1).

2.2.2.2 Summary of reference standard

In all studies, the reference standard was preterm birth, either within 48hrs, and/or within 7 days.[1, 44-64] All 20 included studies evaluated the index tests against the 7-day reference standard.[1, 44-64] Six of the Actim Partus studies (Brik 2010, Goyal 2016, Lembet 2002, Ting 2007, Tripathi 2016 and Vishwekar 2017) and one PartoSure study (Werlen 2015) also evaluated the index test against a 48hr reference standard.[44, 51, 54, 55, 58-60] Quantitative fFN was not evaluated against the 48hr reference standard.[45, 46, 64]

2.2.2.3 Summary of tests administration

The manufacturers' descriptions of the index tests and how they should be used are presented in section 1.3.1, Table 1. The quantity and quality of reported details regarding how each test was performed within a study varied considerably.

2.2.2.3.1 Actim Partus

From the 16 studies which used the Actim Partus test, typically the information provided on how the test was administered followed the manufacturer's guidance.[1, 46, 48-57 58-60] There were, however, the following differences.

The reporting of detection limit thresholds varied between studies, but it is likely that all studies used a threshold of 10µg/l: Eight studies (Abo El-Ezz, Cooper, Danti 2011, Goyal 2016, Lembet 2002, Riboni 2011, Tripathi 2016 and Vishwekar 2017) clearly reported a detection limit of 10µg/ml.[1, 48, 52, 54-56, 59, 60] Two studies (Altinkaya 2009 and Eroglu 2007) reported that samples higher than 30 µg/l give "*a strong positive result*".[49, 53] It is unclear in both these studies whether a weak positive at 10 µg/l would have been considered as positive result although this appears to be the case.[49, 53] One study (Brik 2010) states that a threshold of 30µg/l was required for a positive result, and that this shows as two blue lines on the dipstick, but this is incorrect (two blue lines show at 10µg/l). Finally, five studies (Azlin 2010, APOSTEL-1, Hadzi-Lega 2017, Tanir 2009 and Ting 2007) did not report what detection limit they used. However, given the qualitative nature of the test, it appears most likely that a 10µg/l threshold was used.[44, 46, 47, 50, 57, 58] Indeed, the manufacturer's guidance indicates that a concentration of 10µg/l or more in the cervical fluid causes a positive Actim Partus test reaction result.

All studies report taking their sample around the external cervical orifice or cervical os or a cervical specimen.[1, 46-60] However, two studies (APOSTEL-1 and Ting 2007) report that the sample was taken from the posterior fornix[46, 58]. The instructions from the manufacturer state that the sample should be taken from the cervical os. Our obstetric clinical experts have advised that samples taken from posterior fornix would likely yield a higher false negative rate since the secretion samples differ between the two areas and concentrations are likely

to be weaker from the posterior fornix. A difference in secretion sample concentration between cervical locations has been demonstrated by Kuhrt (2014), although using the quantitative fFN test, it is likely that Actim Partus would be affected in a similar manner.[65]

The manufacturer's instructions state that if the single (control) line does not appear then the test is invalid. However, one study (Tanir 2009) interpreted no visible lines as a positive test result.[57] This was their way of dealing with missing data due to invalid results. This is unlikely to greatly alter results as only two tests from 68 results were invalid.[57] The remaining studies did not report details on how invalid tests were treated.

Two studies (APOSTEL-1 and Cooper 2012) both froze their samples at -20°C for future analysis.[1, 46] With APOSTEL-1 samples were reported to have been transferred and stored at -80°C within 6 months.[46] It is unknown how long after the transfer samples remained in storage before testing, however it is likely that total storage time would have exceeded 6 months.[46] Likewise, in Cooper (2012), it is unclear how long the tests remained frozen before testing.[1] Both of these studies (APOSTEL-1 and Cooper 2012) go on to describe that samples were thawed before the Actim Partus test was run.[1, 46] This protocol differs to the manufacturer's guidance (and, of course, clinical practice) where, freezing a sample is not discussed in the instructions for use. On the manufacturer's website, further information on freezing is provided; Medix Biochemica suggest that storage should be at -20°C for 'longer periods', however they provide little detail on this and stipulate that the effect of thawing on the test performance should be tested by the user (they assume no responsibility or liability if the antibody has been frozen). We have received clinical input from our Obstetricians to suggest that freezing is unlikely to affect the sample's integrity and therefore unlikely to impact test accuracy.

2.2.2.3.2 PartoSure

All four studies assessing the PartoSure test (Bolotskikh 2017, Hadzi-Lega 2017, Nikolova 2015, Werlen 2015) appeared to perform the test in a manner that was consistent with the manufacturer's guidance.[44, 47, 61-63] However, Bolotskikh (2017) did not specifically report that samples were collected using a speculum.[61]

2.2.2.3.3 Fetal Fibronectin

Of the two studies which used the quantitative fFN test, one study (EUIFS) appeared to perform the test in accordance with the manufacturer's guidance.[64] The other study (APOSTEL-1) froze the samples as reported in section 2.2.2.3.1 above.[45] This is not in accordance with the manufacturer's guidance, which states that in order to avoid degradation of the analyte, frozen samples should be assayed within three months.

2.2.2.3.4 Time to test results

Based on manufacturer's instructions, the maximum time between taking the sample and receiving the test results (including time for mixing in solvents etc.) is approximately 6 minutes for Actim Partus and PartoSure and 12 minutes for quantitative fFN. Some of the studies provided information within their methods relating to the maximum amount of time test personnel had to wait before reading the result from the test. This was always in accordance with manufacturer's guidance.

None of the 20 studies included in our review reported, in the study results, how long it took to perform the test and receive a result. No further data, therefore, can be presented on this outcome.

2.2.2.3.5 Test failure rates

Only two studies (Goyal 2016, Tanir 2009), both of which evaluated Actim Partus, reported information about test failures: Goyal 2016 reported that there were no invalid tests and Tanir 2009 reported that there were two cases where the Actim Partus test failed to show any visible lines. For these two women, the test result was not assigned as invalid and the tests were not re-run; they were instead assigned as positive Actim Partus results.[54, 57] It was not clear whether these two women had a preterm delivery within 48 hrs or within 7 days (i.e. whether they were assigned as TP or FP), and therefore it was not possible to conduct sensitivity analyses where these two cases were assigned as negative test results.

2.2.2.3.6 Frozen samples

Following the protocol amendment to include studies using frozen samples, two additional studies (APOSTEL-1 and Cooper 2012) were included.[1, 45] Methodological details on freezing of the samples are described in sections 2.2.2.3.1 and 2.2.2.3.2 and are further discussed in section 2.2.5. We have received clinical input from our Obstetricians to suggest that freezing of the sample is unlikely to affect the integrity of the sample, even if thawed outside of the timeframe suggested by the manufacturer's guidance, and is, therefore, unlikely to have a major impact on test accuracy. However, it should be noted we have no published data to verify this.

2.2.3 Description of included participants

2.2.3.1 Studies evaluating more than one index test

The APOSTEL-1 study, which evaluated both Actim Partus and quantitative fFN in the same population, recruited 350 participants.[45, 46] The characteristics of these participants are reported in Table 4. To summarise, mean maternal age was 29.9 (SD=5.4) years and mean

gestational age at presentation was 29.0 (SD=2.7) weeks. In this study 20% (n=71) of the participants were multiple gestations and 23% (n=79) had previously delivered preterm.[45, 46]

The other study evaluating more than one index test in the same population was a much smaller study; Hadzi-Lega 2017 assessed both Actim Partus and PartoSure and recruited 57 participants.[47] The characteristics of these participants are also reported in Table 4. In this study, median maternal age was reported as 27 (IQR 23.0-30.5) years and median gestational age at presentation was reported as 31 (IQR 28.8-32.4) weeks.[47]

2.2.3.2 Actim Partus

In the 16 Actim Partus studies (including APOSTEL-1 and Hadzi-Lega 2017), sample sizes ranged from n=30 in Vishwekar (2017) to n=468 in Tripathi (2016).[45-47, 59, 60] Reported maternal age ranged from a mean of 24.5 (SD 5.16) years in Altinkaya (2009) to a median of 31(IQR 28-34) years in Danti (2011). Gestational age ranged from a mean of 28.7 (SD not reported) in Riboni (2011) to 32.8 (SD=3.24) weeks in Goyal (2016).[49, 52, 54, 56] Further details describing the participant characteristics in each study are given in Table 4.

Three of the 16 studies assessing Actim Partus (APOSTEL-1, Cooper 2012 and Vishwekar 2017) included participants with multiple gestations (20% ,n=71; 6%, n=20 and 7%, n=2 respectively).[1, 46, 60] Previous preterm delivery was reported in seven studies (APOSTEL-1, Brik 2010, Cooper 2012, Eroglu 2007, Goyal 2016, Lembet 2002, Vishwekar 2017) and ranged from 3.9% (n=2) in Eroglu (2007) to 30% (n=18) in Goyal (2016).[1, 46, 51, 53-55, 60]

2.2.3.3 PartoSure

In the four PartoSure studies (Hadzi-Lega 2017, Bolotskikh 2017, Nikolova 2015 and Werlen 2015) sample sizes ranged from n=41 in Werlen (2015) to n=203 in Nikolova (2015).[44, 47, 61-63] Reported maternal age ranged from a median of 25 (IQR 23-38) years in Bolotskikh (2017) to a median of 27 years in Nikolova (2015) and Hadzi-Lega (2017) (range 18-43 in Nikolova 2015; IQR 23-30.5 in Hadzi-Lega 2017).[47, 61-63] Mean gestational age ranged from 29.5 (SD 2.91) years in Werlen (2015) to a median of 32 weeks (range 20.5-36.6) in Nikolova (2015).[44, 62, 63] Further details describing the participant characteristics in each study are given in Table 4.

2.2.3.4 Quantitative fFN

In the two studies assessing quantitative fFN (APOSTEL-1 and EUIFS), sample sizes were n=350 and n=455 respectively.[45, 64] Maternal age was similar in these two studies (mean 29.9 years, SD 5.4 in APOSTEL-1 and mean 29.5 years, SD 5.2 in EUIFS), as was gestational age (mean 29.0 weeks, SD 2.7 in APOSTEL-1 and median 29.6 weeks, IQR= 26.7-31.6 weeks in EUIFS).[45, 64] The APOSTEL-1 study had a higher proportion of multiple

pregnancies than EUIFS (20%, n=71 versus 15%, n=67).[45, 64] Also, proportionally more women had previously delivered preterm in APOSTEL-1 than in EUIFS (23%, n=79 versus 16%, n=72).[45, 64] Further details describing the participant characteristics in each study are given in Table 4.

Table 4 Participant Characteristics

		Participants *n*	Maternal Age mean ± SD	Gestational Age at Presentation (weeks) mean ± SD	Multiple Gestations *n* (%)	BMI kg/m^2 mean ± SD	Gravidity mean ± SD	Parity mean ± SD	Previous Preterm Delivery *n(%)*	Previous Miscarriage/ Stillbirth *n(%)*	Mode of delivery *n(%)*
Study assessing Actim Partus and quantitative fFN											
APOSTEL-1[45, 46]		350	29.9 ± 5.4	29.0 ± 2.7	71 (20)	23.1 ± 4.3	NR	NR	79 (23)	NR	Non-spontaneous deliveries within reference standard timeframe excluded
Study assessing Actim Partus and PartoSure											
Hadzi-Lega, 2017[47]		57	27(23.0-30.5)[A]	31 (28.8 - 32.4)[A]	0 (0)	NR	NR	NR	NR	NR	Non-spontaneous deliveries within reference standard timeframe excluded
Actim Partus											
Abo El-Ezz, 2014[48]		57	27.40 ± 6.1	29.70 ± 2.5	0 (0)	NR	NR	2.91 ±NR	NR	NR	NR
Altinkaya, 2009[49]		105 [E]	24.52 ± 5.16	29.63 ± 4.4	0 (0)	24.1 ± 3.5	NR	0.65 ± 0.95	0 (0)	NR	NR
Azlin, 2010 [50]	phIGFBP-1 (+)	7	29.57 ± 3.99	32.96 ± 3.07 [C]	0 (0)	NR	2.43 ± 1.27	1.00 ± 1.16	NR	0.43 ± 1.13	Spontaneous:36 (70.6) Elective caesarean: 7 (13.7)
	phIGFBP-1 (-)	44	28.34 ± 4.32	32.38 ± 2.64 [C]	0 (0)	NR	2.59 ± 1.59	0.91 ± 0.96	NR	0.68 ± 1.25	Emergency caesarean: 4 (7.8) Unknown: 4 (7.8)
Brik, 2010[51]		276	29.4 ± 5.9 (15-46) [D]	29.9 ± 2.8 (23-34) [D]	0 (0)	NR	NR	Null = 58.3%	26 (9.4)	NR	Unclear
Cooper, 2012[1]		349	29 ± 5.0 (17-46) [D]	29+6 (4+6); [24-34][A]	20 (5.7)[*F]	NR	NR	Null = 43.3%	56 (16.1)	NR	Spontaneous: 182 (52.1) Operative: 52 (14.9) Caesarean: 115 (33.0)
Danti, 2011 [52]		60 [G]	31 (28-34)[A]	30.0 (28.7-31.4)[A]	0 (0)	NR	NR	Null= 63%	NR	NR	NR
Eroglu, 2007[53]		51 [E]	27.6 ± 3.5	29.5 ± 2.6	0 (0)	22.6 ± 2.9	NR	0.4 ± 0.6	2 (3.9)	2 (3.9)	NR
Goyal, 2016[54]		60	29.92 ± 5.14	32.84 ± 3.24	0 (0)	23.61 ± 2.45	NR	0.9 ± 0.3	18 (30)	NR	NR
Lembet, 2002[55]		36 [E]	28.4 ± 5.3	31.3 ± 3.3	0 (0)	<19.6kg/m^2 =7 >26kg/m^2=29	2.2 ± 1.6	0.7 ± 1.1	7 (16)	3 (7)	Unclear
Riboni, 2011[56]	Delivering term	210	30.4 ± 5.6	28.7 ± NR	0 (0)	23.36 ± 3.99	NR	NR	NR	NR	NR
	Delivering preterm		30.7 ± 5.1			22.72 ± 3.7					

Tanir, 2009[57]	phIGFBP-1 (+)[H]	25	28.4 ± 4.6	30.6 ± 3.5	0 (0)	25.1 ± 3.5	2.1 ± 1.3	0.7 ± 0.3	NR	1.8 ±0.8	Caesarean: 18 (72.0) [J] Vaginal: 16 (64)[J]
	phIGFBP-1 (-)[I]	43	28.4 ± 5.3	29.6 ± 2.3	0 (0)	26.9 ± 4.4	2.2 ± 1.3	0.6 ± 0.4	NR	1.5 ± 0.5	Caesarean: 20 (46) [J] Vaginal: 14 (33) [J]
Ting, 2007[58]	phIGFBP-1 (+)[H]	28	27	30.5	0 (0)	NR	2	0.5	NR	NR	NR
	phIGFBP-1 (-)[I]	66	27	32.2	0 (0)	NR	2	1	NR	NR	
Tripathi, 2016[59]		468	NR	NR	NR	NR	NR	NR	0 (0)	NR	NR
Vishwekar 2017[60]	phIGFBP-1 (+)	14	25 (19-35)	32	2 (6.7)	'normal limits'	NR		4 (13.3)	NR	Unclear
	phIGFBP-1 (-)	16		32.5							
PartoSure											
Bolotskikh, 2017[61]		99(100)	25(23-38)[A]	32 (29-36)[A]	0 (0)	NR	NR	Null=32%	15 (15)	27 (27)	Non-spontaneous deliveries within reference standard timeframe excluded
Nikolova, 2015[62, 63]		203	27 (18-43)[B]	32.0 (20.5-36.6)[B]	0 (0)	NR	NR	NR	NR	NR	Patients with non-spontaneous delivery excluded (n=8), unclear when these occurred
Werlen, 2015[44]		41	27.6 ± 5.3 (18-39) [D]	29.5 ± 2.91 (24-34)	0 (0)	NR	NR	0.54 ± 0.71	NR	NR	NR
Quantitative fFN											
EUIFS [64]		455	29.5 ± 5.2	29.6 (26.7-31.6)[A]	67 (15)	(n=429) 24.5 (22.0-28.0)[A]	NR	Null=55%	72 (16)	NR	Non-spontaneous deliveries within reference standard timeframe excluded

Notes: **A** Data are given as median (interquartile range); **B** Data are given as median (range); **C** Reported as POA; **D** Data are given as mean ± SD (range); **E** Study group (symptomatic) **F** 2 participants, number of foetuses unknown **G** High risk group: CL≤ 30 mm; **H** Actim Partus test positive; **I** Actim Partus test negative; **J** Data as reported in paper, however values do not add up.

Key: BMI, body mass index; fFN, fetal fibronectin; n, Number; Null, nulliparous; NR, not reported; SD, standard deviation

2.2.4 Summary of any treatments given

The index tests of interest are designed to be used before the reference standard (the occurrence of preterm delivery within 48hrs and/or 7 days). As such, it is important to consider the use of any treatments that might impact upon the reference standard and thus the test accuracy results. Whether a woman received treatment for symptoms of preterm labour varied substantially between studies. This variability was based on i) the standard treatment protocols used within each study and ii) the test(s) used to initiate treatment within the study (e.g. Actim Partus, PartoSure or quantitative fFN or another test that is not being assessed as part of this review e.g. transvaginal cervical length). Subsequently, where treatment was given, the number of women receiving treatment was not always reported, particularly with reference to the results of the diagnostic tests of interest. This means that, in several cases, it is difficult to ascertain the extent to which treatments may have impacted upon the test accuracy results.

2.2.4.1 Studies evaluating more than one index test

Treatment decisions from the APOSTEL-1 trial were not based on the test results of Actim Partus or quantitative fFN alone. Instead, treatment decisions were based on the combined results of two tests not being assessed in this review (cervical length measurement and quantitative fFN with a threshold of 50ng/ml). Additionally, a strict treatment protocol was not used; instead recommendations were provided. Tocolytics (a choice from nifedipine, indomethacin, atosiban and ritodrine) were recommended for women with a cervical length <10mm but not for women >30mm. Women with a cervical length between 10 and 30mm and a positive fFN result were encouraged to receive tocolytics, whereas those with a negative fFN received tocolytics at the discretion of the advising clinician. Corticosteroids were permitted at the clinician's discretion. No data were provided on how many women received tocolytics or corticosteroids.[45, 46]

Treatment decisions from the Hadzi-Lega (2017) study (a single centre study assessing both Actim Partus and PartoSure) followed the standard of care at the hospital. This included hospitalization, discharge, tocolytics (including beta mimetics and calcium channel blockers) and corticosteroids (e.g. betamethasone). Thirty-eight of 57 women (67%) received corticosteroids, four of whom had a preterm birth within 7 days. Also, thirty-eight of 57 women (67%) received tocolytics (note: not all patients who received corticosteroids received tocolytics), six of whom had a preterm birth within 7 days. No details were provided for either of the test results in correlation to treatment administration.[47]

2.2.4.2 Actim Partus

Sixteen studies assessed Actim Partus, two of which (APOSTEL-1 and Hadzi-Lega 2017) also assessed another index test and have been discussed above (section 2.2.4.1).[45-47] The remaining 14 studies (Abo El-Ezz 2014, Altinkaya 2009, Azlin 2010, Brik 2010, Cooper 2012, Danti 2011, Eroglu 2007, Goyal, 2016, Lembet 2002, Riboni 2011, Tanir 2009, Ting 2007, Tripathi 2016, Vishwekar 2017) all report the use of tocolytics and most specify the type (typically atosiban, calcium channel blockers, nifedipine, ritodrine or magnesium sulphate).[1, 48-57 58-60]. All of these studies except Azlin (2010) report the use of corticosteroids (either betamethasone or dexamethasone where specified).[50]

One study reports that cervical length measurements were available to managing clinicians and were used to aid treatment decisions (Danti 2011).[52] Of the 60 patients in this study 22 (37%) received tocolytics and 28 (47%) received corticosteroids.

In another study (Eroglu 2007) it is reported that patients were admitted to hospital based on frequency of contractions and digital examination. In this study women with a gestational age >34 weeks did not receive any tocolytics or corticosteroids. In total 16 out of 51 patients (31%) received tocolytics, eight of whom had a positive Actim Partus result; the study does not report the number of patients that received corticosteroids.[53] In a further four studies (Lembet 2002, Riboni 2011, Tripathi 2016, Vishwekar 2017) treatment decisions were also guided by gestational age.[55, 56, 59, 60] In Lembet (2002) patients >34 weeks gestation were not administered tocolytics, women >24 weeks gestation received corticosteroids and those <28 weeks received a repeat dose. In this study 21 out of 36 patients (58%) received tocolytics, eight of whom had a positive Actim Partus test result.[55] In Riboni (2011) women less than <34 weeks gestation received corticosteroids, while all patients received tocolytics.[56] In both Tripathi (2016) and Vishwekar (2017) women <34 weeks gestational age received tocolytics and corticosteroids.[59, 60] The studies by Riboni (2011), Tripathi (2016) and Vishwekar (2017) do not report the proportion of patients that received these treatments (or the Actim Partus results of these patients).[56, 59, 60]

Five studies state that treatment decisions were based on standard hospital protocol (Azlin 2010, Brik 2010, Cooper 2012, Goyal 2016, and Ting 2007).[1, 50, 51, 54, 58] In these studies it is largely unclear whether diagnostic tests, symptoms or maternal characteristics (e.g. gestational age) were used to guide decision making. Four of these studies report details on the proportion of patients receiving treatment: in Azlin (2010) 12 out of 51 patients (24%) received tocolytics (2 of whom had a positive Actim Partus result), but no details were reported regarding how many patients received corticosteroids;[50] in Brik (2010) 213 out of 276 patients (77%) received tocolytics while 200 (73%) received corticosteroids;[51] in Cooper (2012), 8 out of 349 patients received tocolytics (2%) and 56 (16%) received

corticosteroids;[1] and in Goyal (2016) all patients received tocolytics while the number receiving corticosteroids was unclear.[54] The number of patients receiving tocolytics or corticosteroids was not reported in Ting (2007).[58]

The remaining 3 studies report limited detail regarding how treatment decisions were made (Abo El-Ezz 2014, Altinkaya 2009 and Tanir 2009).[48, 49, 57] In Abo El-Ezz (2014) and Altinkaya (2009) all patients were admitted to hospital.[48, 49] Abo El-Ezz (2014) report that all patients received tocolytics, while it is unclear how many patients received corticosteroids.[48] In Altinkaya (2009) patients whose symptoms persisted received tocolytics, while all patients were administered corticosteroids.[49] Neither of these studies report details regarding the Actim Partus results of those receiving treatment. In Tanir (2009), 63 out of 69 patients (93%) received tocolytic therapy (23 of whom had a positive Actim Partus test result). The number of patients receiving corticosteroids was not reported.[57]

2.2.4.3 PartoSure

Four studies assessed PartoSure, one of which (Hadzi-Lega 2017) also assessed another index test and has been discussed above (section 2.2.4.1).[47]

In one of the remaining three studies (Bolotskikh 2017), tocolytics were given irrespective of test outcomes (all admitted women were treated).[61] In another study (Werlen 2015) treatment was at the discretion of the investigator according to the protocol of the department. Of the 41 women in this study, 13 received corticosteroids and 25 tocolytics, but no details were provided for treatment administration in connection with the PartoSure test results.[44] In the remaining study (Nikolova 2015) no treatment protocols were reported.[62, 63]

2.2.4.4 Quantitative fFN

Two studies assessed quantitative fFN, one of which (APOSTEL-1) also assessed another index test and has been discussed above (section 2.2.4.1).[45-47]

In the other study (EUIFS), treatment with tocolytics and steroids were based on a combination of cervical length and the quantitative fFN result at a threshold of 50ng/ml, neither of which are index tests in this review. In addition, no data were provided on how many women received treatments.

2.2.5 Quality appraisal of included studies

Quality appraisal was conducted, using Phase 3 of the QUADAS-2 tool,[42] for all 20 studies. Phase 3 of the QUADAS-2 tool contains four domains: patient selection, index tests, reference standard, and flow and timing. The APOSTEL-1 study and the study by Hadzi-Lega (2017) each assess two index tests (Actim Partus and quantitative fFN; PartoSure and

Actim Partus, respectively).[45-47] For these two studies, the index test domain was conducted separately for each test.

It is important to note that the QUADAS-2 tool assesses the likely risk of bias and not the presence or magnitude of bias. Any rating of 'high', therefore means that the *risk* of bias is high but does not mean that there is a high degree of bias, or even that bias has been detected. The quality of the included studies is discussed in the sections that follow, and a summary of the QUADAS-2 ratings is provided in Table 5.

Table 5 Quality Appraisal, QUADAS-2[42]

		AP & fFN[b]	***AP & PS[c]***	***Actim Partus***														***PartoSure***			***fFN***
		APOSTEL-1	*Hadzi-Lega, 2017*	*Abo El-Ezz, 2014*	*Altinkaya, 2009*	*Azlin, 2010*	*Brik, 2010*	*Cooper, 2012*	*Danti, 2011*	*Eroglu, 2007*	*Goyal, 2016*	*Lembet, 2002*	*Riboni, 2011*	*Tanir, 2009*	*Ting, 2007*	*Tripathi, 2016*	*Vishwekar 2017*	*Bolotskikh, 2017*	*Nikolova, 2015*	*Werlen, 2015*	*EUIFS*
Patient selection	Was a consecutive or random sample of patients enrolled?	U	U	U	U	Y	Y	U	Y	U	U	U	U	U	U	U	Y	U	Y	U	U
	Was a case-control design avoided?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Did the study avoid inappropriate exclusions?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Could the selection of patients have introduced bias?	U	U	U	U	L	L	U	L	U	U	U	U	U	U	U	L	U	L	U	U
	Is there concern that the included patients do not match the review question?	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Index test 1	Were the index test results interpreted without knowledge of the results of the reference standard?	U	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	If a threshold was used, was it pre-specified?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Could the conduct or interpretation of the index test have introduced bias?	U	L	L	L	L	L	U	L	L	L	L	L	L	L	L	L	L	L	L	L
	Is there concern that the index test, its conduct, or interpretation differ from the review question?	H[a]	L	L	L	L	L	L	L	L	L	L	L	L	H[a]	L	L	L	L	L	L
Index test 2	Were the index test results interpreted without knowledge of the results of the reference standard?	U	Y	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	If a threshold was used, was it pre-specified?	Y	Y	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	Could the conduct or interpretation of the index test have introduced bias?	U	L	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	Is there concern that the index test, its conduct, or interpretation differ from the review question?	L	L	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Reference standard	Is the reference standard likely to correctly classify the target condition?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Were the reference standard results interpreted without knowledge of the results of the index test?	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	Could the reference standard, its conduct, or its interpretation have introduced bias?	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
	Is there concern that the target condition as defined by the reference standard does not match the review question?	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Flow and timing	Was there an appropriate interval between index test(s) and reference standard?	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Did all patients receive a reference standard?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Did patients receive the same reference standard?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Were all patients included in the analysis?	N	N	N	U	Y	N	Y	Y	Y	N	Y	U	Y	N	N	Y	Y	N	N	N
Could the patient flow have introduced bias?	H	H	H	U	L	H	H	L	L	H	L	U	L	H	H	L	L	H	H	H

Key: AP, Actim Partus; H, high; L, low; N, no; n/a, not applicable; PS, PartoSure U, unknown; Y, yes

Notes: a, The samples were taken from the posterior fornix of the vagina rather than the external cervical os; b, index test 1 = Actim Partus, index test 2 = fFN' c, index test 1 = Actim Partus, index test 2 = PartoSure.

2.2.5.1 Patient selection

All included studies were single-gate DTA studies, and thus avoided the use of a case-control design (i.e. there were no studies that selected a group of women who had delivered pre-term and a group of control group women who did not deliver preterm)[1, 44-64]. In addition, all included studies avoided inappropriate exclusion of participants in terms of their participant inclusion and exclusion criteria. However, only five of the 20 included studies (Azlin, 2010; Brik, 2010; Danti, 2011; Nikolova, 2015; Vishwekar, 2017) were rated as having a low risk of bias due to patient selection (four assessing Actim Partus and one assessing PartoSure).[50-52, 60, 62, 63] These studies reported that eligible women were enrolled into the study consecutively.[50-52, 60, 62, 63] For the remaining 15 studies, it was unclear whether patient selection could have introduced bias because it was unclear whether a consecutive or random sample of participants was recruited.

For all studies included in the review, there were no concerns about whether or not the included participants matched the *review* question (Table 5). However, it should be reiterated that in two studies (APOSTEL-1, Danti 2011), all women included in final analyses of index test data had a transvaginal cervical length measurement ≤30mm, which would likely increase the prevalence of preterm birth in these studies.[45, 46, 52]

2.2.5.2 Index tests

Two of the included studies enable a direct comparison between two of the index tests of interest, by evaluating both index tests in the same population; the APOSTEL-1 study evaluated both Actim Partus and quantitative fFN and the study by Hadzi-Lega (2017) evaluated both Actim Partus and PartoSure.[45-47] The other 18 studies all evaluated only one of the index tests of interest, with three of the studies (Bolotskikh, 2017; Nikolova, 2015; Werlen 2015) only evaluating PartoSure,[44, 61-63] the EUIFs study only evaluating quantitative fFN (Bruijn, 2016),[64] and the remaining 14 studies only evaluating Actim Partus.

All studies included in the review, except for the APOSTEL-1 study and the study by Cooper (2012),[1, 45, 46] were rated as low risk with regards to whether the conduct and interpretation of the test could have introduced bias (Table 5).This is because all studies either clearly reported pre-specified thresholds for the test (quantitative fFN) or used a test with a standardised threshold (Actim Partus or PartoSure), and due to the timing of the tests, all studies other than the APOSTEL-1 and the Cooper (2012) study interpreted the index tests without knowledge of the reference standard (the tests were conducted before the occurrence of preterm birth).[1, 45, 46] In both the APOSTEL-1 study and the Cooper (2012) study, frozen samples were used (see section 2.2.2.3.6).[1, 45, 46] In both studies, whilst samples were collected prior to the assessment of the reference standard, the index tests

(Actim Partus and quantitative fFN in APOSTEL-1 and Actim Partus in Cooper 2012) were interpreted after the assessment of the reference standard.[1, 45, 46] For these index tests, it is unclear as to whether these interpretations were made blind to whether preterm birth had occurred.[1, 45, 46] It should be noted, however, that this is unlikely to lead to a high risk of bias for either test because there is limited need for any subjective interpretation of the test results.

Although not covered by QUADAS-2, it should also be considered whether, in the studies assessing more than one test (APOSTEL-1 and Hadzi-Lega, 2017)[45-47] and in those that included a clinical assessment or a test not included in this review (e.g. transvaginal cervical length), whether clinicians were blinded to this information when interpreting the index test. This is to mitigate any 'cross-contamination' of test results (i.e. bias in the interpretation of the test due to prior knowledge from another test or other clinical information).

Even though, as previously mentioned, there is limited scope for bias to occur in the interpretation of any of the index test results, it should be noted that the quantitative fFN test does not require any subjective judgement of the test result, whereas both the Actim Partus and PartoSure require some judgement (albeit limited). In the use of the Actim Partus test, the potential for bias when interpreting results is still greater than for the quantitative fFN test where no subjective interpretation is required. Due to this, in the APOSTEL-1 study, Actim Partus was performed before the quantitative fFN test.[45, 46] However, additional tests were also performed in the APOSTEL-1 study (qualitative fFN and cervical length measurement) and these were conducted before the index tests; it is unclear whether the index tests were interpreted blind to the results from the qualitative fFN test or the cervical length measurement. It is therefore unclear whether any 'cross-contamination' of results may have occurred between the two tests, although again, due to the nature of the tests the scope for such bias is very limited. In the study by Hadzi-Lega, 2017, where both the PartoSure and Actim Partus tests were performed, it was unclear which test was performed first, or indeed whether the tests were performed in a predetermined order.[47] The authors do state that the reader of the index tests was blind to the results of ultrasound and digital examinations, and this would mitigate any 'cross-contamination', however limited, between these assessments and the index test results.

In four of the remaining studies, three assessing Actim Partus (Azlin 2010, Goyal 2016, Cooper 2012) and one assessing PartoSure (Bolotskikh 2017), it was reported that the index test was interpreted blind to the results of transvaginal cervical length measurements (Azlin 2010, Bolotskikh 2017, Goyal 2016) or qualitative fFN (Cooper 2012).[1, 50, 54, 61] In the studies by Azlin 2010, Bolotskikh 2017 and Goyal 2016, this was because the cervical length measurement was conducted after the index test and in Cooper 2012 it was stated that the

Actim Partus test was conducted blind to the qualitative fFN measurement. In an additional study (Brik 2010) it was stated that the sample was collected before the cervical length measurement and that the Actim Partus was conducted immediately, so it is likely that the index test was interpreted before the cervical length measurement.[51] In a further seven studies (EUIFS, Danti 2011, Eroglu 2007, Nikolova 2015, Riboni 2011, Ting 2007, Tripathi 2016) additional tests or assessments (qualitative fFN and/or cervical length measurement) were reported, but it was unclear whether the index tests were interpreted blind to the results of these additional tests.[52, 53, 56, 58, 59, 62-64] In the remaining six studies (Abo El-Ezz 2014, Altinkaya 2009, Lembet 2002, Tanir 2009, Vishwekar 2017, Werlen 2016) there was no reported use of other tests or clinical assessments.[44, 48, 49, 55, 57, 60]

In all of the studies included in the review, except for the APOSTEL-1 study and Ting (2007),[45, 46, 58] there were no concerns that the conduct or interpretation of either of the index tests was different from the review question (Table 5). However, in APOSTEL-1 and Ting (2007), the test was conducted in a way that may have impacted the results because samples were taken from the posterior fornix of the vagina rather than the external cervical os.[45, 46, 58]

2.2.5.3 Reference standard

In all of the included studies the reference standard was whether or not preterm birth occurred in 48hrs and/or within 7 days.[1, 44-64] As such, the reference standard and the target condition were identical. Similarly, there were no concerns (in any of the studies) that the target condition, as defined by the reference standard (pre-term birth), did not match the review question Table 5.

In addition, because the reference standard was the occurrence of pre-term birth rather than a diagnostic test, no assessment was made regarding whether the reference standard results were *interpreted* without knowledge of the results of the index test, i.e. because the reference standard was not something that involved interpretation. As such, there were no concerns in any of the included studies regarding whether the reference standard, its conduct, or its interpretation could have introduced bias (Table 5).

2.2.5.4 Flow and timing

Of course, in all of the included studies, all participants 'received' a reference standard.[1, 44-64] Also, because the reference standard was the occurrence or otherwise of preterm delivery by 48hrs and/or by 7 days, this was the same for all participants in all of the included studies (Table 5).

It is important to note that because the index tests in this review (PartoSure, Actim Partus and quantitative fFN) are designed to be conducted before the occurrence of the reference standard, this in itself may introduce bias (through the use of treatments to prevent the occurrence of the reference standard). It is likely that use of tocolytics, though indicated, may inflate false positive results and deflate true positive results.

For all but two of the included studies (APOSTEL-1 and Cooper, 2012), the timing of the index test was as per the manufacturers' instructions.[1, 45, 46] For both of these studies (APOSTEL-1 and Cooper, 2012) the timing of the index tests may have introduced bias because frozen samples were used and it is unclear when the samples were thawed and used.[1, 45, 46] However, this is unlikely to have a large impact upon the test results (see section 2.2.2.3.6 for further details).

Related to this, and outside of the core QUADAS-2 questions, we also assessed blinding of clinical staff to the results of the index test. Awareness of test results may necessarily influence treatment decisions, but may also lead to unintentional differences between those with positive and negative test results in the way in which a patient is managed. On the other hand, when treatment decisions are influenced by a test other than the index test, it may not be possible to ascertain from the literature whether index test positives and negatives received different patterns of clinical management. In fact, only two of the included studies reported that the managing clinicians were aware of the test results (Azlin 2010, Riboni 2011) although in the latter study (Riboni 2011), it was stated that clinical management was not altered by this knowledge.[50, 56] In six of the studies (APOSTEL-1, EUIFS, Cooper 2012, Danti 2011, Hadzi-Lega 2017, Nikolova 2015), it was reported that clinical management personnel were unaware of the index tests (management was based on qualitative fFN results and/or cervical length).[1, 45-47, 52, 62-64] In four of the studies, clinical management personnel were blinded to the results of all tests reported in the study (Eroglu 2007, Ting 2007, Lembet 2002, Tanir 2009) and, therefore, it was unclear on what basis clinical decisions were made.[53, 55, 57, 58] In the remaining eight studies (Abo El-Ezz 2014, Altinkaya 2009, Bolotskikh 2017, Brik 2010, Goyal 2016, Tripathi 2016, Vishwekar 2017, Werlen 2015) it was not clearly reported whether clinical personnel were blinded to the index test results.[44, 48, 49, 51, 54, 59-61]

With regards to missing data, it was clear that in ten of the included studies (APOSTEL-1, Hadzi-Lega 2017, Abo El-Ezz 2014, Brik 2010, Goyal 2016, Ting 2007, Tripathi 2016, Nikolova 2015, Werlen 2015, and the EUIFS study) some of the participants were excluded from the analysis (Table 5).[44-48, 51, 54, 58, 59, 62-64]. This may lead to bias because it was not clear from the study reports whether the women who were not included differed systematically from those whose data were analysed. In fact, only eight studies (Azlin 2010, Cooper 2012,

Danti 2011, Eroglu 2007, Lembet 2002, Tanir 2009, Vishwekar 2017, Bolotskikh 2017) clearly specify that data were analysed from all participants who received tests, although it should be noted that in the study by Tanir (2009), two participants with failed tests were coded as test positives.[1, 50, 52, 53, 55, 57, 60, 61]

2.2.5.5 Quality appraisal summary

All of the included studies were single-gate DTA studies rather than case-control studies. A key issue to note is that only two studies (APOSTEL-1 and Hadzi-Lega 2017) evaluated more than one index test in the same population.[45-47] Thus, only these two studies allow for a direct comparison between the index tests. Any comparisons made between the tests based on the other studies will be subject to confounding (i.e. due to differences in the recruited populations, hospitals, and other factors that may impact upon whether a woman is likely to have a pre-term delivery).

Although all of the included studies recruited appropriate populations, in the majority of studies there was a lack of clarity regarding recruitment procedures and so an assessment cannot be made as to whether the selection of women in these studies could have introduced bias. Indeed, only five studies, four of which were Actim Partus studies (Azlin 2010, Brik 2010, Danti 2011, Vishwekar 2017) and one of which was a PartoSure study (Nikolova 2015), clearly reported selection procedures.[50-52, 60, 62, 63]

In almost all of the studies, index tests were performed in a manner that is consistent with clinical practice. However, both the APOSTEL-1 study and the study by Cooper (2012) used frozen samples, which means that the index tests (Actim Partus in both studies and quantitative fFN in APOSTEL-1) were interpreted after the reference standard.[1, 45, 46] Due to the nature of the tests this is unlikely to have much impact, although some amount of bias cannot be completely ruled out. Additionally, in APOSTEL-1 and Ting (2007), samples were taken from the posterior fornix of the vagina rather than the external cervical os. This sampling method is not compatible with the manufacturer's guidance for the Actim Partus test.[45, 46, 58]

Furthermore, although not likely to have a major impact upon results, in several studies there is potential for 'cross-contamination' of results from one index test to another and/or from additional tests that are not part of this review (e.g. cervical length measurement). There is also potential, in several studies, for the index test results to have influenced the clinical management of patients: approximately half of the studies stated that the clinicians involved in patient management were unaware of the index test results (APOSTEL-1, EUIFS, Cooper 2012, Danti 2011, Hadzi-Lega 2017, Nikolova 2015, Eroglu 2007, Ting 2007, Lembet 2002, Tanir 2009).[1, 45-47, 52, 53, 55, 57, 58, 62-64]

Another point to note is that the index tests are designed to be used before the occurrence of the reference standard, and this was the case in all but two studies (APOSTEL-1 and Cooper 2012).[1, 45, 46] This may inflate false positive results and deflate true positive results through the use of tocolytics to prevent the occurrence of the reference standard.

Ten of the studies (APOSTEL-1, Hadzi-Lega 2017, Abo El-Ezz 2014, Brik 2010, Goyal 2016, Ting 2007, Tripathi 2016, Nikolova 2015, Werlen 2015, and the EUIFS study) were rated as being of high risk of bias with regards to not including all women in analyses.[44-48, 51, 54, 58, 59, 62-64] High risk of bias does not equate to a high degree of bias, and this should be considered here and for other QUADAS-2 items where a high risk of bias rating has been given (Table 5).

2.2.6 Results of quantitative data synthesis (test accuracy data)

2.2.6.1 Seven day delivery reference standard

2.2.6.1.1 Studies evaluating more than one index test

Two studies (APOSTEL-1 and Hadzi-Lega 2017) reported test accuracy data on two index tests.[45-47] Both studies only used the 7 day (and not the 48hr) delivery reference standard. Prevalence of preterm birth within 7 days was 19.7% (95% CI 15.7, 24.3) in the APOSTEL-1 study and 10.5% (95% CI 4.0, 21.5) in the Hadzi-Lega 2017 study.

The APOSTEL-1 study reports test accuracy from 350 women for both Actim Partus and quantitative fFN (Table 6).[45, 46] The sensitivity and specificity for Actim Partus, against delivery within 7 days, were 78.3% (95% CI 66.7, 87.3) and 89.3% (95% CI 85.1, 92.7) respectively. As would be expected, in the quantitative fFN results from the APOSTEL-1 study, lowering the threshold for a positive test result increased sensitivity and decreased specificity whereas elevating the threshold for a positive test result increased specificity and decreased sensitivity (see Table 6). The quantitative fFN sensitivity and specificity values that were most similar to Actim Partus values for the APOSTEL-1 study were those provided at a threshold of 200 ng/ml, where sensitivity was 71.0% (95% CI 58.8-83.1) and specificity 83.6% (95% CI 78.8, 87.8). With regard to the quantitative fFN data from APOSTEL-1, the threshold with the highest PPV was 500ng/ml (70.7%; 95%CI 54.5, 83.9) and the lowest was at 10ng/ml (28.9%; 95% CI 23.2, 35.3). For Actim Partus the PPV of delivery within 7 days was 64.3% (95% CI 53.1, 74.4). For quantitative fFN, the highest NPV was at a threshold of 10ng/ml (97.5%; 95% CI 93.0, 99.5) and the lowest at a threshold of 500ng/ml (87.1%; 82.8, 90.6). For Actim Partus, NVP was 94.4% (95% CI 90.9, 96.8). Likelihood ratios, concordance and yield were also calculated for this study and these values are provided in Table 6.

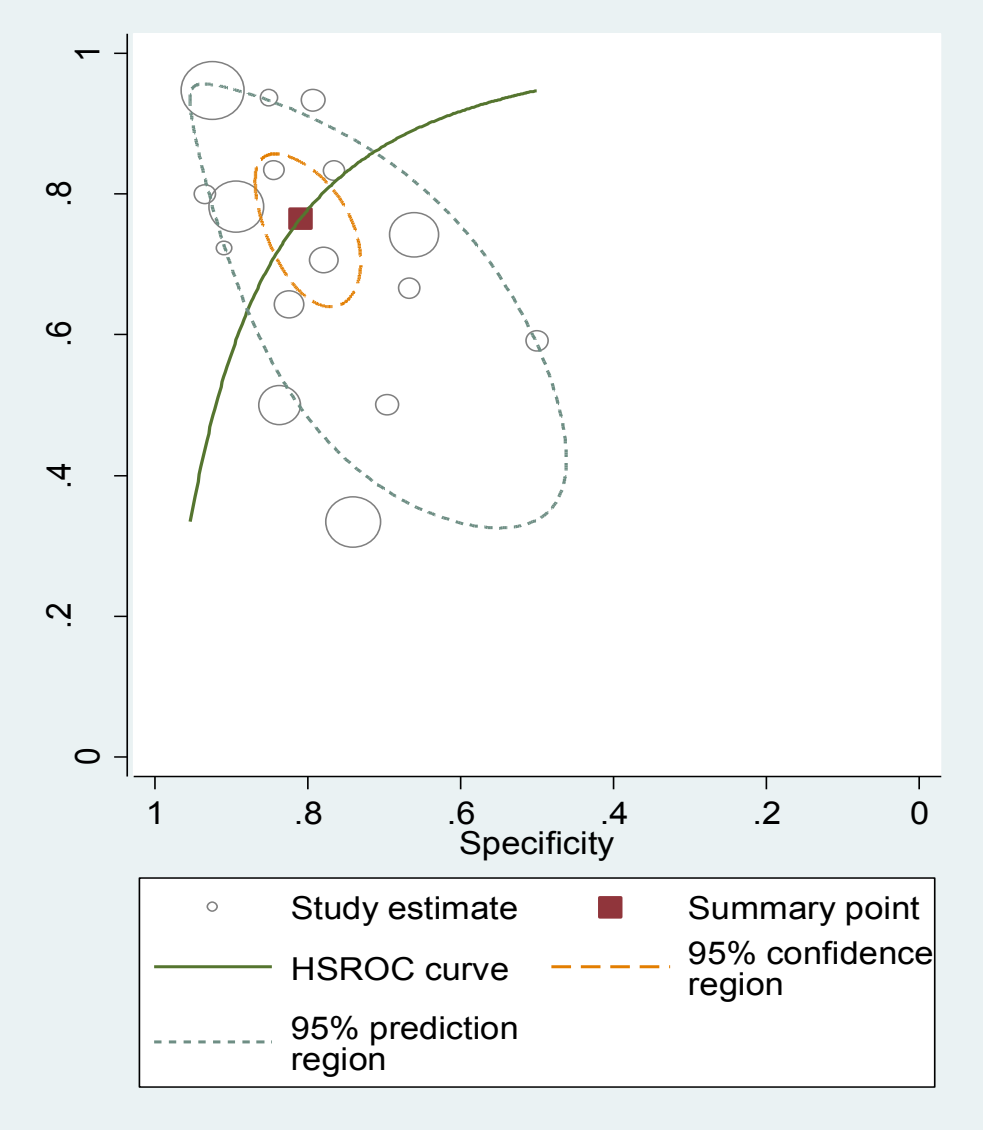
The other study providing data on more than one index test in the same sample was Hadzi-Lega (2017), where test accuracy data from 57 women were reported for both Actim Partus and PartoSure (Table 6).[47] The sensitivity of both Actim Partus and PartoSure for delivery within 7 days was 83.3% (95% CI 35.9, 99.6) whilst specificity was higher for PartoSure 90.2% (95% CI 78.6-96.7) compared to Actim Partus 76.5% (95% CI 62.5, 87.2). In addition, PPV, LR+ and concordance were higher for PartoSure than for Actim Partus, although the wide confidence intervals, particularly for PPV, are notable. LR- and NVP were similar for both tests and diagnostic yield was higher for Actim Partus than PartoSure. Specific values are given in Table 6.

2.2.6.1.2 Actim Partus

Results for Actim Partus against the 7-day delivery reference standard were provided by 16 studies (Table 6, Figure 4). Across these studies, sensitivity ranged from 33.3% (95% CI 4.3, 77.7) in Cooper (2012) to 94.7% (95% CI 89.9, 97.7) in Tripathi (2016). Specificity of Actim Partus ranged from 50.0% (95% CI 24.7-75.3) in Goyal (2016) to 93.5% (95% CI 82.1-98.6) in Azlin (2010). The three studies with the lowest sensitivity were Cooper 2012 (33.3%; 95% CI 4.3, 77.7), Danti 2011 (50%; 95% CI 6.8, 93.2) and Riboni (50%; 95% CI 15.7, 84.3). The prevalence of these studies was much lower (prevalence ranging from 1.7, 95% CI 0.6, 3.7 to 6.7, 95% CI 1.8, 16.2) compared to all other studies (prevalence ranging from 9.8, 95% CI 3.3-21.4 to 73.3, 95% CI 60.3, 83.9). Indeed, the large range of prevalence estimates across these studies is particularly noteworthy (see section 2.2.2.1.1). Meanwhile, the three studies with the lowest specificities were Goyal 2016 (50%; 95% CI 24.7, 75.3), Brik 2010 (66%; 95% CI 59.8, 72.0), and Abo El-Ezz 2014 (67%; 95% CI 46.0, 83.5). There were no obvious methodological or participant characteristics in these studies to explain the differences, and although two of these studies had high prevalence (Goyal, 2016; Abo El-Ezz 2014) the other (Brik, 2010) did not.

A summary ROC plot for all 16 studies assessing Actim Partus against the 7-day delivery reference standard is provided in Figure 4. Pooled analyses were performed for these data and provided a pooled sensitivity of 0.77 (95% CI 0.68, 0.83) and a pooled specificity of 0.81 (95% CI 0.76, 0.85).

Figure 4 ROC plot for Actim Partus against the 7-day reference standard



Key: HSROC, Hierarchical summary receiver-operating characteristic

Data from these 16 Actim Partus studies were also used to calculate LR+, LR -, PPV, NPV, concordance and yield. These values are provided in Table 6.

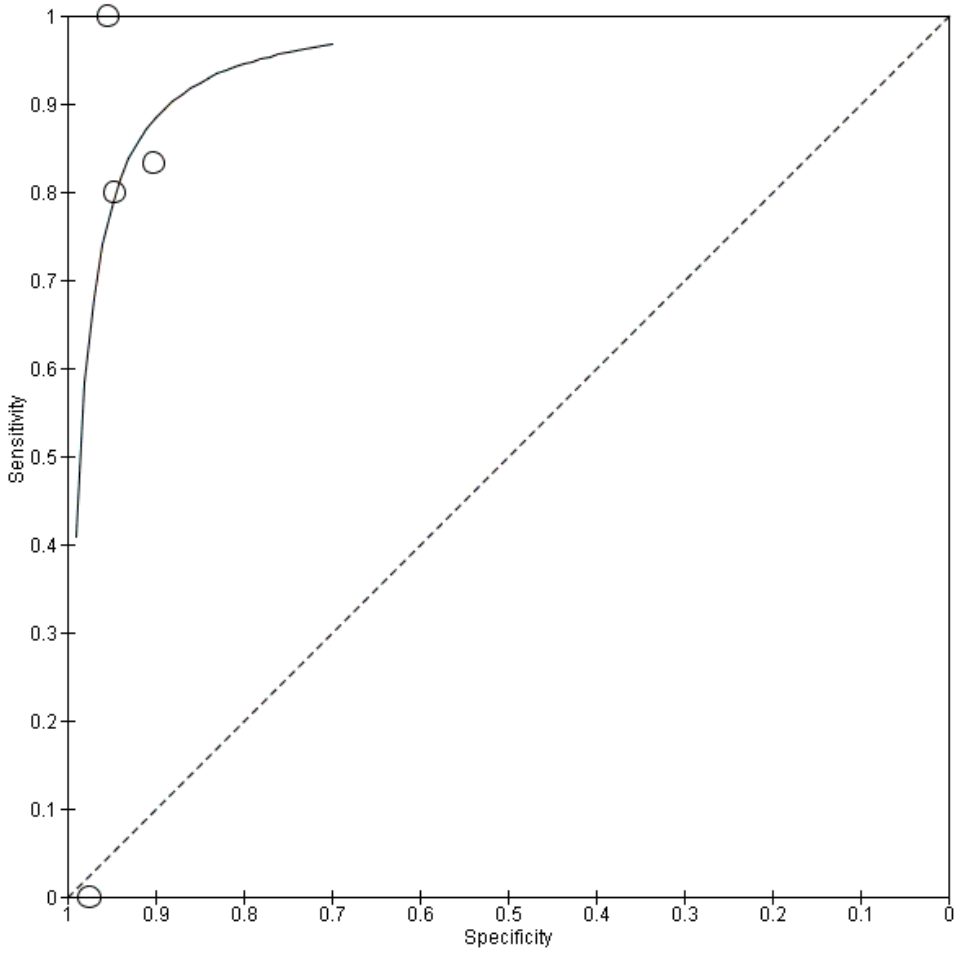
2.2.6.1.3 PartoSure

Results for PartoSure against the 7-day delivery reference standard were reported in four studies (Table 6, Figure 5). Prevalence of preterm delivery within 7 days ranged from 2.4% (95% CI 0.1, 12.9) in Werlen (2015) to 17.2% (95% CI 12.3, 23.2) in Nikolova (2015). These four studies had wide ranging sensitivity, from 0% (95% CI 0.0, 97.5) in Werlen (2015) to 100% (95% CI 73.5-100.0) in Bolotskikh (2017), whereas specificity was more similar across studies, ranging from 90.2 (95% CI 78.6, 96.7) in Hadzi-Lega (2017) to 97.5% (95% CI 96.8, 99.9) in Werlen (2015). The low sensitivity, from Werlen (2015), was due to the fact that in the sample of 41 participants, only one tested (falsely) positive using the PartoSure test.

Discounting this study, the sensitivity range would be 80% (95% CI 63.1, 91.6) in Nikolova (2015) to 100% (95% CI 73.5-100.0) in Bolotskikh (2017).

A summary ROC plot for the four studies assessing PartoSure against the 7-day delivery reference standard is provided in Figure 4. Pooled analyses were performed for these data and provided a pooled sensitivity of 0.83 (95% CI 0.61, 0.94) and a pooled specificity of 0.95 (95% CI 0.89, 0.98).

Figure 5 ROC plot for PartoSure against the 7-day reference standard



Data from these four PartoSure studies were also used to calculate LR+, LR -, PPV, NPV, concordance and yield. These values are provided in Table 6.

2.2.6.1.4 Quantitative fFN

Results against the 7 day delivery reference standard for the two quantitative fFN studies (EUIFS and APOSTEL-1), at the three thresholds (10ng/ml, 200ng/ml and 500ng/ml), are presented in Table 6. Prevalence of pre-term birth within the sample was lower in EUIFS (10.5%; 95% CI 7.9, 13.7) compared to APOSTEL-1 (19.7%; 95% CI 15.7, 24.3). EUIFS presented with slightly lower (within 2%) sensitivity values compared to APOSTEL-1 at both

the 10 and 200ng/ml threshold (Table 6). At the 500ng/ml threshold the sensitivity was much lower (29.2%; 95% CI 17.0, 44.1 in EUIFS vs 42%; 95% CI 30.2, 54.4) in APOSTEL-1. Similarly, specificity values were slightly lower (within 5%) in EUIFS compared with APOSTEL-1 at both the 200ng/ml and 500ng/ml thresholds, whereas at the 10ng/ml threshold the specificity was much lower (32.2%; 95% CI 27.7, 37.0 in EUIFS vs 42.3%; 95% CI 36.5, 48.4 in APOSTEL-1).

Data from the two quantitative fFN studies were also used to calculate LR+, LR -, PPV, NPV, concordance and yield. These values are provided in Table 6.

Table 6 Calculated diagnostic accuracy parameters against the 7 day reference standard

Study		Participants (N)	Sensitivity %	Specificity %	LR+	LR-	PPV %	NPV %	Prevalence %	Concordance	Yield
Study assessing quantitative fFN and Actim Partus											
APOSTEL-1 [45, 46]	fFN at 10ng/ml	350	95.7; 87.8-99.1	42.3; 36.5-48.4	1.66; 1.48-1.86	0.10; 0.03-0.31	28.9; 23.2-35.3	97.5; 93.0-99.5	19.7; 15.7-24.3	0.53; 0.48-0.58	0.65; 0.60-0.70
	fFN at 200ng/ml	350	71.0; 58.8-81.3	83.6; 78.8-87.8	4.34; 3.20-5.88	0.35; 0.24-0.50	51.6; 41.1-62.0	92.2; 88.1-95.1	19.7; 15.7-24.3	0.81; 0.77-0.85	0.27; 0.23-0.32
	fFN at 500 ng/ml	350	42.0; 30.2-54.5	95.7; 92.7-97.8	9.84; 5.30-18.28	0.61; 0.49-0.74	70.7; 54.5-83.9	87.1; 82.8-90.6	19.7; 15.7-24.3	0.85; 0.81-0.89	0.12; 0.09-0.16
	Actim Partus	350	78.3; 66.7-87.3	89.3; 85.1-92.7	7.33; 5.11-10.51	0.24; 0.16-0.38	64.3; 53.1-74.4	94.4; 90.9-96.8	19.7; 15.7-24.3	0.87; 0.83-0.91	0.24; 0.20-0.29
Study assessing PartoSure and Actim Partus											
Hadzi-Lega, 2017 [47]	PartoSure	57	83.3; 35.9-99.6	90.2; 78.6-96.7	8.50; 3.43-21.03	0.18; 0.03-1.11	50.0; 18.7-81.3	97.9; 88.7-99.9	10.5; 4.0-21.5	0.90; 0.79-0.96	0.18; 0.09-0.30
	Actim Partus	57	83.3; 35.9-99.6	76.5; 62.5-87.2	3.54; 1.92-6.52	0.22; 0.04-1.31	29.3; 10.3-56.0	97.5; 86.8-99.9	10.5; 4.0-21.5	0.77; 0.64-0.87	0.30; 0.18-0.43
Actim Partus											
Abo El-Ezz, 2014 [48]	Actim Partus	57	66.7; 47.2-82.7	66.7; 46.0-83.5	2.00; 1.11-3.61	0.50; 0.28-0.89	69.0; 49.2-84.7	64.3; 44.1-81.4	52.6; 39.0-66.0	0.67; 0.53-0.79	0.51; 0.37-0.64
Altinkaya, 2009 [49]	Actim Partus	105	64.3; 35.1-87.2	82.4; 73.0-89.6	3.66; 2.02-6.61	0.43; 0.21-0.88	36.0; 18.0-57.5	93.8; 86.0-97.9	13.3; 7.5-21.4	0.80; 0.71-0.87	0.24; 0.16-0.33
Azlin, 2010 [50]	Actim Partus	51	80.0; 28.4-99.5	93.5; 82.1-98.6	12.27; 3.77-39.86	0.21; 0.04-1.24	57.1; 18.4-90.1	97.7; 88.0-99.9	9.8; 3.3-21.4	0.92; 0.81-0.98	0.14; 0.06-0.26
Brik, 2010 [51]	Actim Partus	276	74.2; 55.4-88.1	66.1; 59.8-72.0	2.19; 1.67-2.87	0.39; 0.21-0.71	21.7; 14.3-30.8	95.3; 90.9-97.9	11.2; 7.8-15.6	0.67; 0.61-0.73	0.38; 0.33-0.44
Cooper, 2012 [1]	Actim Partus	349	33.3; 4.3-77.7	74.1; 69.1-78.6	1.28; 0.41-4.04	0.90; 0.51-1.59	2.2; 0.3-7.7	98.4; 96.1-99.6	1.7; 0.6-3.7	0.73; 0.68-0.78	0.26; 0.22-0.31
Danti, 2011 [52]	Actim Partus	60	50.0; 6.8-93.2	69.6; 55.9-81.2	1.65; 0.57-4.74	0.72; 0.27-1.94	10.5; 1.3-33.1	95.1; 83.5-99.4	6.7; 1.8-16.2	0.68; 0.55-0.80	0.32; 0.20-0.45
Eroglu, 2007 [53]	Actim Partus	51	83.3; 35.9-99.6	84.4; 70.5-93.5	5.36; 2.48-11.56	0.20; 0.03-1.19	41.7; 15.2-72.3	97.4; 86.5-99.9	11.8; 4.4-23.9	0.84; 0.71-0.93	0.24; 0.13-0.38
Goyal, 2016 [54]	Actim Partus	60	59.1; 43.2-73.7	50.0; 24.7-75.3	1.18; 0.68-2.04	0.82; 0.45-1.50	76.5; 58.8-89.3	30.8; 14.3-51.8	73.3; 60.3-83.9	0.57; 0.43-0.69	0.57; 0.43-0.69
Lembet, 2002 [55]	Actim Partus	36	93.8; 69.8-99.8	85.0; 62.1-96.8	6.25; 2.19-17.88	0.07; 0.01-0.49	83.3; 58.6-96.4	94.4; 72.7-99.9	44.4; 27.9-61.9	0.89; 0.74-0.97	0.50; 0.33-0.67
Riboni, 2011 [56]	Actim Partus	210	50.0; 15.7-84.3	83.7; 77.8-88.5	3.06; 1.43-6.54	0.60; 0.30-1.20	10.8; 3.0-25.4	97.7; 94.2-99.4	3.8; 1.7-7.4	0.82; 0.77-0.87	0.18; 0.13-0.24
Tanir, 2009 [57]	Actim Partus	68	93.3; 68.1-99.8	79.2; 65.9-89.2	4.50; 2.61-7.74	0.08; 0.01-0.56	56.0; 34.9-75.6	97.7; 87.7-99.9	22.1; 12.9-33.8	0.82; 0.71-0.91	0.37; 0.25-0.49
Ting, 2007 [58]	Actim Partus	94	70.6; 44.0-89.7	77.9; 67.0-87.6	3.20; 1.90-5.38	0.38; 0.18-0.80	41.4; 23.5-61.1	92.3; 83.0-97.5	18.1; 10.9-27.4	0.77; 0.67-0.85	0.31; 0.22-0.41
Tripathi 2016, [59]	Actim Partus	468	94.7; 89.9-97.7	92.4; 88.9-95.1	12.43; 8.45-18.30	0.06; 0.03-0.11	85.7; 79.5-90.6	97.3; 94.8-98.8	32.5; 28.3-37.0	0.93; 0.91-0.95	0.36; 0.32-0.41
Vishwekar, 2017 [60]	Actim Partus	30	72.2; 46.5-90.3	90.9; 58.7-99.8	7.94; 1.20-52.62	0.31; 0.14-0.66	92.9; 66.1-99.8	66.7; 38.4-88.2	62.1; 42.3-79.3	0.79; 0.60-0.92	0.48; 0.29-0.67

Study		Participants (N)	Sensitivity %	Specificity %	LR+	LR-	PPV %	NPV %	Prevalence %	Concordance	Yield
PartoSure											
Bolotskikh, 2017 [61]	PartoSure	99	100.0; 73.5-100.0	95.4; 88.6-98.7	21.75; 8.35-56.64	0.00; NA	75.0; 47.6-92.7	100.0; 95.7-100.0	12.1; 6.4-20.2	0.96; 0.90-0.99	0.16; 0.10-0.25
Nikolova, 2015 [62, 63]	PartoSure	203	80.0; 63.1-91.6	94.6; 90.1-97.5	14.93; 7.74-28.80	0.21; 0.11-0.41	75.7; 58.8-88.2	95.8; 91.5-98.3	17.2; 12.3-23.2	0.92; 0.88-0.95	0.18; 0.13-0.24
Werlen, 2015 [44]	PartoSure	41	0.0; 0.0-97.5	97.5; 96.8-99.9	0.00; NA	1.03; 0.98-1.08	0.0; 0.0-97.5	97.5; 96.8-99.9	2.4; 0.1-12.9	0.95; 0.84-0.99	0.02; 0.00-0.13
Quantitative fFN											
EUIFS [64]	fFN at 10ng/ml	455	93.8; 82.8-98.7	32.2; 27.7-37.0	1.38; 1.25-1.53	0.19; 0.06-0.59	14.0; 10.4-18.3	97.8; 93.6-99.5	10.5; 7.9-13.7	0.39; 0.34-0.43	0.71; 0.66-0.75
	fFN at 200ng/ml	455	70.8; 55.9-83.0	78.6; 74.3-82.5	3.31; 2.55-4.30	0.37; 0.24-0.58	28.1; 20.3-37.0	95.8; 93.1-97.7	10.5; 7.9-13.7	0.78; 0.74-0.82	0.27; 0.23-0.31
	fFN at 500 ng/ml	455	29.2; 17.0-44.1	94.3; 91.6-96.4	5.16; 2.85-9.34	0.75; 0.63-0.90	37.8; 22.5-55.2	91.9; 88.8-94.3	10.5; 7.9-13.7	0.99; 0.84-0.90	0.08; 0.06-0.11

Key: LR+, Positive likelihood ratio; LR- Negative likelihood ratio; PPV positive predictive value; NPV negative predictive value

2.2.6.2 48 hour delivery reference standard

Seven of the included studies also provided test accuracy data for the index tests against a 48hr preterm delivery reference standard. Six of these studies evaluated Actim Partus (Brik 2010, Goyal 2016, Lembet 2002, Ting 2007, Tripathi 2016, Vishwekar 2017) and one evaluated PartoSure (Werlen 2015).[44, 51, 54, 55, 58-60]

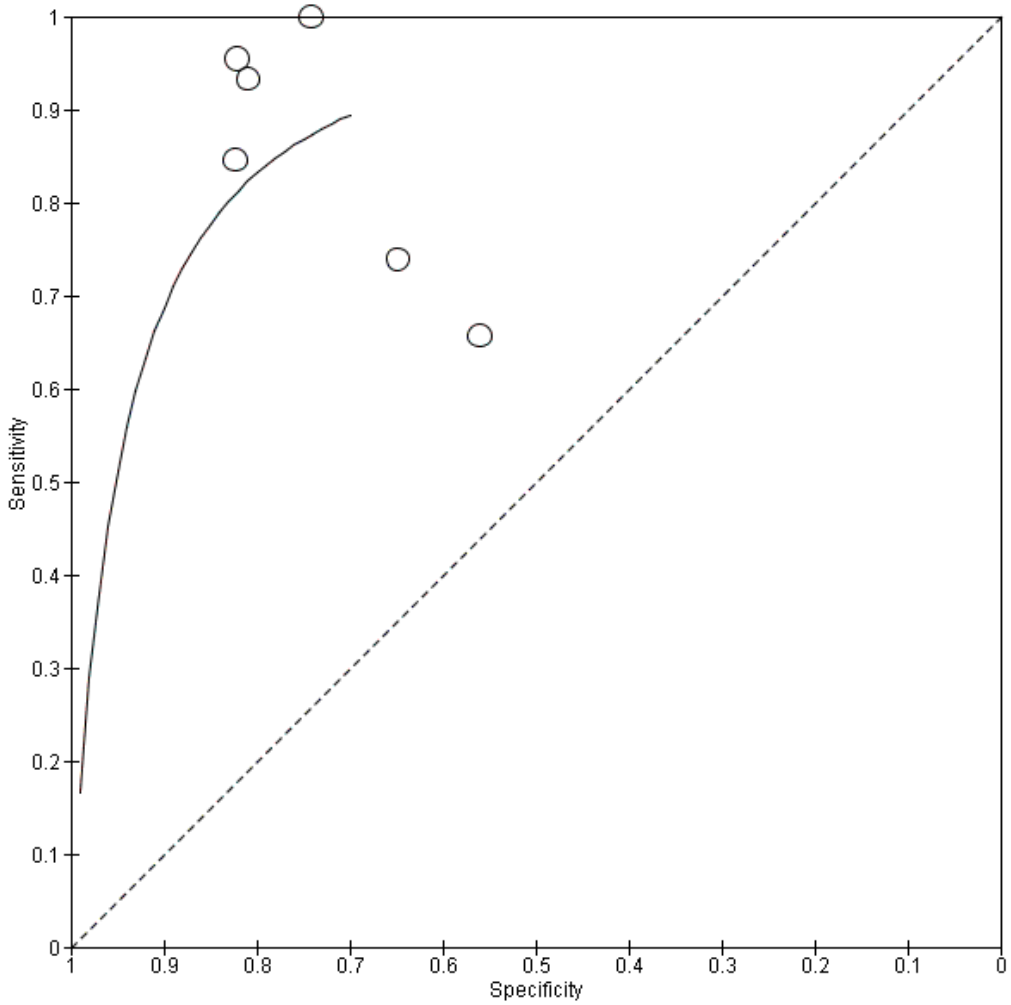
2.2.6.2.1 Actim Partus

Across the six studies evaluating Actim Partus against the occurrence of preterm birth within 48hrs (Brik 2010, Goyal 2016, Lembet 2002, Ting 2007, Tripathi 2016, Vishwekar 2017), prevalence of preterm birth within 48hrs ranged from 5.3% (95% CI 1.7, 12.0) in Ting (2007) to 58.3% (95% CI 44.9, 70.9) in Goyal (2016). Sensitivity ranged from 65.7% (95% CI to 47.8, 80.9) in Goyal (2016) to 100.0% (95% CI 47.8, 100.0) in Ting (2007). Specificity ranged from 56.0% (95% CI 34.9, 75.6) in Goyal (2016) to 82.4% (95% CI 56.6, 96.2) in Vishwekar (2017).

Specific sensitivity and specificity values for all six studies are given in Table 7, where it can be seen that the sensitivity and specificity of Actim Partus for the 48hr reference standard were lowest in Goyal (2016) and Brik (2010) compared to the other four studies (Lembet 2002, Ting 2007, Tripathi 2016 and Vishwekar 2017), which seem more in line with each other. There were no obvious methodological or participant characteristic differences in these studies (other than the women in Goyal 2016 and Brik 2010 were on average a year older than those in the other studies). Brik (2010) did present data for the number of women who received tocolytics (77.2%), but the other five studies did not provide this information, so we cannot assess whether this was particularly high or low in comparison.

A ROC plot for the six studies assessing Actim Partus against the 48 hour delivery reference standard is provided in Figure 6. Pooled analyses were performed for these data and provided a pooled sensitivity of 0.87 (95% CI 0.74, 0.94) and a pooled specificity of 0.73 (95% CI 0.62, 0.82).

Figure 6 ROC plot for Actim Partus against the 48 hour delivery reference standard



In the six studies evaluating Actim Partus against the occurrence of preterm birth within 48hrs, the PPV was lower for Brik (2010) and Ting (2007) (16.0%, 95% CI 9.6, 24.4 and 17.9%, 95% CI 6.1, 36.9 respectively) compared to the other four studies (range 62.1%, 95% CI 54.4, 69.5 to 78.6%, 95% CI 49.2-95.3). This is likely to be linked to the prevalence also being low in these two studies (8.3%, 95% CI 5.4, 12.2 in Brik 2010 and 5.3%, 95% CI 1.7-12.0 in Ting 2007). Conversely, NPV was lowest for Goyal 2016 (53.8%, 95% CI 33.4, 73.4), with NVP in the other five studies ranging from 87.5% (95% CI 61.7, 98.4) in Vishwekar (2017) to 100% (95% CI 94.6-100.0) in Ting (2007). Looking at the diagnostic yield, Goyal (2016) was the only study where over 50% of the population had a positive Actim Partus result (57%; 95% CI 0.43-0.69) all other studies had a diagnostic yield 50% or less. Data from these six Actim Partus studies were also used to calculate LR+, LR -, and concordance. These values are provided in Table 7.

2.2.6.2.2 PartoSure

In the study evaluating PartoSure against the occurrence of preterm birth within 48hrs (Werlen 2015), prevalence of preterm birth within 48hrs was lower (2.4%; 95% CI 0.1, 12.9) than in any of the Actim Partus studies discussed in section 2.2.6.2.1. Sensitivity was 0.0% (95% CI 0.0, 97.5) and specificity 97.5% (95% CI 86.8, 99.9); the total sample size was 41 and only one test result was positive (a false positive).[44] These data, along with calculated values for PPV, NVP, LR+, LR-, concordance and yield are given in Table 7.

Table 7 Calculated diagnostic accuracy parameters against the 48 hour delivery reference standard

Study	Participants (N)	Sensitivity	Specificity	LR+	LR-	PPV	NPV	Prevalence	Concordance	Yield
Actim Partus										
Brik, 2010 [51]	276	73.9%; 51.6-89.8	64.8%; 58.6-70.7	2.10; 1.56-2.82	0.40; 0.20-0.81	16.0%; 9.6-24.4	96.5%; 92.5-98.7	8.3%; 5.4-12.2	0.66; 0.60-0.71	0.38; 0.33-0.44
Goyal, 2016 [54]	60	65.7%; 47.8-80.9	56.0%; 34.9-75.6	1.49; 0.90-2.47	0.61; 0.34-1.09	67.6%; 49.5-82.6	53.8%; 33.4-73.4	58.3%; 44.9-70.9	0.62; 0.48-0.74	0.57; 0.43-0.69
Lembet, 2002 [55]	36	93.3%; 68.1-99.8	81.0%; 58.1-94.6	4.90; 2.01-11.96	0.08; 0.01-0.55	77.8%; 52.4-93.6	94.4%; 72.7-99.9	41.7%; 25.5-59.2	0.86; 0.71-0.95	0.50; 0.33-0.67
Ting, 2007 [58]	94	100.0%; 47.8-100.0	74.2%; 63.8-82.9	3.87; 2.72-5.50	0.00; NA	17.9%; 6.1-36.9	100.0%; 94.6-100.0	5.3%; 1.7-12.0	0.76; 0.66-0.84	0.30; 0.21-0.40
Tripathi, 2016 [59]	468	95.5%; 89.7-98.5	82.1%; 77.8-86.0	5.34; 4.26-6.69	0.06; 0.02-0.13	62.1%; 54.4-69.5	98.3%; 96.1-99.5	23.5%; 19.7-27.6	0.85; 0.82-0.88	0.36; 0.32-0.41
Vishwekar, 2017[60]	30	84.6%; 54.6-98.1	82.4%; 56.6-96.2	4.79; 1.67-13.74	0.19; 0.05-0.68	78.6%; 49.2-95.3	87.5%; 61.7-98.4	43.3%; 25.5-62.6	0.83; 0.65-0.94	0.47; 0.28-0.66
PartoSure										
Werlen, 2015 [44]	41	0.0%; 0.0-97.5	97.5%; 86.8-99.9	0.00; NA	1.03; 0.98-1.08	0.0%; 0.0-97.5	97.5%; 86.8-99.9	2.4%; 0.1-12.9	0.95; 0.84-0.99	0.02; 0.00-0.13

Key: LR+, Positive likelihood ratio; LR- Negative likelihood ratio; PPV positive predictive value; NPV negative predictive value

2.3 Results of the overview

In this section we present an overview of studies that assess the diagnostic test accuracy of at least one of the index tests (PartoSure, Actim Partus, and quantitative fFN), in addition to a qualitative fFN test and/or quantitative fFN at 50ng/ml test. In a similar manner, a further overview incorporating cervical length is given in Appendix 3.

The reasons for presenting these additional data are twofold. First, fFN at 50ng/ml is recommended in the NICE guidelines for current practice.[26] Second, these tests are also comparators in our review of clinical effectiveness (end-to-end studies) but no such studies were found (see Chapter 3). However, it is important to highlight here that, because fFN at 50ng/ml was not an index test in the test accuracy review, the data are presented for information only and do not form part of the systematic review of test accuracy. Nevertheless, only data presented in those studies that were included in the systematic review are included in this overview. Due to this, these data are not exhaustive of all data available for qualitative fFN or quantitative fFN at 50ng/ml (because several studies only reporting such data would have been excluded from the systematic review). To ensure that the data presented here are similar to the wider available evidence, we also identified, for comparison, recent systematic reviews of test accuracy data for qualitative fFN or quantitative fFN at 50ng/ml.

In this overview, diagnostic test accuracy data, against a reference standard of preterm birth within 7 days, are provided for fFN, at a threshold of 50ng/ml (qualitative or quantitative test). These data are presented in Table 8, together with test accuracy data for index tests from the same studies (i.e. data from index tests produced for the systematic review of PartoSure, Actim Partus, and quantitative fFN at thresholds other than 50ng/ml, see section 2.2).

Table 8 Test accuracy results (against preterm birth within 7 days) for index tests and fFN at 50ng/ml

Study	Test	Participants (*n*)	Sensitivity %; 95%CI	Specificity %; 95%CI	PPV %; 95%CI	NPV %; 95%CI
APOSTEL-1 [45, 46]	fFN@10	350	95.7; 87.8-99.1	42.3; 36.5-48.4	28.9; 23.2-35.3	97.5; 93.0-99.5
	fFN@50[a]	350	91.3; 82.0-96.7	64.8; 58.9-70.3	38.9; 31.3-46.9	96.8; 93.2-98.8
	fFN@200	350	71.0; 58.8-81.3	83.6; 78.8-87.8	51.6; 41.1-62.0	92.2; 88.1-95.1
	fFN@500	350	42.0; 30.2-54.5	95.7; 92.7-97.8	70.7; 54.5-83.9	87.1; 82.8-90.6
	Actim Partus	350	78.3; 66.7-87.3	89.3; 85.1-92.7	64.3;53.1-74.4	94.4; 90.9-96.8
EUIFS [64]	fFN@10	455	93.8; 82.8-98.7	32.2; 27.7-37.0	14.0; 10.4-18.3	97.8; 93.6-99.5
	fFN@50[a]	455	89.6; 77.3-96.5	62.2; 57.3-66.9	21.8; 16.3-28.3	98.1; 95.5-99.4
	fFN@200	455	70.8; 55.9-83.0	78.6; 74.3-82.5	28.1; 20.3-37.0	95.8; 93.1-97.7

	fFN@500	455	29.2; 17.0-44.1	94.3; 91.6-96.4	37.8; 22.5-55.2	91.9; 88.8-94.3
Cooper, 2012 [1]	fFN@50[d]	291	33.3; 4.3-77.7	89.8; 85.7-93.1	6.5; 0.8-21.4	98.5; 96.1-99.6
	Actim Partus	349	33.3; 4.3-77.7	74.1; 69.1-78.6	2.2; 0.3-7.7	98.4; 96.1-99.6
Eroglu, 2007 [53]	fFN@50[b]	51	83.3; 35.9-99.6	80.0; 65.4-90.4	35.7; 12.8-64.9	97.3; 85.8-99.9
	Actim Partus	51	83.3; 35.9-99.6	84.4; 70.5-93.5	41.7; 15.2-72.3	97.4; 86.5-99.9
Nikolova, 2015 [62, 63]	fFN@50[b]	66	50.0; 21.1-78.9	72.2; 58.4-83.5	28.6; 11.3-52.2	86.7; 73.2-94.9
	PartoSure	203	80.0; 63.1-91.6	94.6; 90.1-97.5	75.7; 58.8-88.2	95.8; 91.5-98.3
Riboni, 2011[56]	fFN@50[c]	210	50.0; 15.7-84.3	80.2; 74.0-85.5	9.1; 2.5-21.7	97.6; 93.9-99.3
	Actim Partus	210	50.0; 15.7-84.3	83.7; 77.8-88.5	10.8; 3.0-25.4	97.7; 94.2-99.4
Ting, 2007[58]	fFN@50[d]	94	56.3; 29.9-80.2	75.6; 64.6-84.7	32.1; 15.9-52.4	89.4; 79.4-95.6
	Actim Partus	94	70.6; 44.0-89.7	77.9; 67.0-86.6	41.4; 23.5-61.1	92.3; 83.0-97.5
Tripathi, 2016 [59]	fFN@50[b]	468	23.8; 17.3-31.4	99.1; 97.3-99.8	92.3; 79.1-98.4	73.2; 68.7-77.3
	Actim Partus	467	94.7; 89.9-97.7	92.4; 88.9-95.1	85.7; 79.5-90.6	97.3; 94.8-98.8

Key: fFN, fetal fibronectin; n, number; NPV, negative predictive value; PPV, positive predictive value
Notes: a, quantitative Rapid fFN 10Q Cassette; b, QuikCheck fFN; c, fFN measured by ELISA; d, fFN testing method unclear,

2.3.1 Test accuracy data for fFN at 50ng/ml

2.3.1.1 Quantity and quality of the data available for fFN at 50ng/ml

As can be seen in Table 8, eight of the 20 included studies (APOSTEL-1, EUIFS, Cooper 2012, Eroglu 2007, Nikolova 2015, Riboni 2011, Ting 2007 and Tripathi 2016) report diagnostic test accuracy data for fFN measured at 50ng/ml threshold (in addition to data for at least one index test).[1, 45, 46, 53, 56, 58, 59, 62-64]

Two studies (APOSTEL-1 and EUIFS) used the quantitative fFN test and report data at 50ng/ml.[45, 46, 64] The APOSTEL-1 study did additionally use a qualitative version of fFN (Rapid fFN for TLiIQ system test) but data for this test are not provided in the included papers.[45, 46]Three studies (Eroglu 2007, Nikolova 2015, Tripathi 2016) used the QuikCheck version of the qualitative fFN test.[53, 59, 62, 63] One further study (Riboni 2011) used the ELISA laboratory technique.[56] The remaining two studies (Cooper 2012 and Ting 2007) did not report which test was used.[1, 58] More specifically, Ting (2007) only states that the test used was a bedside test that was 'qualitatively reported'.[58] Cooper (2012) reports using a fFN test manufactured by Adeza Biochemical Corporation,[1] but as this company produces both an ELISA testing method and the Rapid fFN for TLiIQ system test,[66] it remains unclear which test was used.

As previously mentioned, in the APOSTEL-1 study, samples for the quantitative fFN test (at 50ng/ml) were collected and frozen for later analysis (this was not the case for the Rapid fFN

for TLiIQ system test). This could potentially introduce a risk of bias as investigators may have known the outcome of the reference standard when interpreting the test.[45, 46] However, as previously noted, due to the nature of the test, the potential for interpretation bias is minimal. For all other studies reporting fFN data at 50ng/ml, the tests were conducted and results analysed at the point of admission, before the reference standard of delivery within 7 days had occurred. However, for three of the studies (Eroglu 2007, Riboni 2011 and Tripathi 2016) it is unclear whether assessors were aware of the Actim Partus test results when analysing fFN tests.[53, 56, 59] Again, although 'cross-contamination' between tests cannot be completely ruled out, the potential for such bias in these types of test is minimal.

One study (Riboni 2011) uses the ELISA technique to determine fFN status (and possibly Cooper (2012) as well, although this is unclear).[1, 56] ELISA is a quantitative technique which was used in a qualitative capacity using 50ng/ml as the threshold, this is the standard threshold, suggesting that in this study the threshold was established *a priori*.[56] However, neither Riboni (2011) nor Cooper (2012) explicitly report pre-specification of the threshold for this test.[1, 56] Of course the APOSTEL-1 study also uses a quantitative fFN test, but multiple pre-specified thresholds were used.[45, 46]

An additional consideration in these 50ng/ml fFN data is that one study (Nikolova 2015) reports fFN at 50ng/ml accuracy data for only 66 out of the 203 patients recruited and included in the PartoSure analyses. The reasons for this are unclear.[62, 63]

2.3.1.2 Test accuracy of 50ng/ml threshold for fFN

Diagnostic test accuracy data for fFN at a threshold of 50ng/ml (against the 7 day delivery reference standard) are provided in Table 8. Sensitivity of fFN at the 50ng/ml threshold ranged from 23.8% (95% CI 17.3, 31.4) in Tripathi (2016) to 91.3% (95% CI 82.0, 96.7) in APOSTEL-1 (quantitative fFN data at 50ng/ml).[45, 46, 59] Specificity ranged from 62.2% (95% CI 57.3, 66.9) in EUIFS to 99.1% (95% CI 97.3, 99.8) in Tripathi (2016).[59, 64] Values for PPV and NPV were also calculated and are presented in Table 8.

Again, it should be noted that these data do not cover all available evidence regarding test accuracy of fFN at a threshold of 50ng/ml and are based only on data reported by studies included in our systematic review of Actim Partus, PartoSure and quantitative fFN at thresholds other than 50ng/ml.

2.3.1.3 Comparison of fFN at 50ng/ml test and index tests

In six studies (APOSTEL-1, Cooper 2012, Eroglu 2007, Riboni 2011, Ting 2007, Tripathi 2016), both fFN at a threshold of 50ng/ml and Actim Partus were assessed in the same sample.[1, 45, 46, 53, 56, 58, 59] One study (Nikolova 2015) assessed fFN at a threshold of 50ng/ml

and PartoSure in the same sample,[62, 63] and two studies (APOSTEL-1, EUIFS) assessed fFN at a threshold of 50ng/ml and fFN at other thresholds in the same sample.[45, 46, 64] Note that the APOSTEL-1 study assessed more than one index test, in addition to fFN at 50ng/ml, in the same sample.[45, 46]

When compared with Actim Partus, sensitivity (against the preterm birth within 7 days reference standard) was higher for fFN at 50ng/ml in one study (APOSTEL-1),[45, 46] lower for fFN at 50ng/ml in two studies(Ting 2007, Tripathi 2016),[58, 59] and the same for both tests in three studies (Cooper 2012, Eroglu 2007 and Riboni 2011).[1, 53, 56] Specificity (against the preterm birth within 7 days reference standard) was lower for fFN compared to Actim Partus in four of the six studies (APOSTEL-1, Eroglu 2007, Riboni 2011 and Ting 2007),[45, 46, 53, 56, 58] and higher for the other two studies (Cooper 2012 and Tripathi 2016).[1, 59] These data are presented in Table 8.

In the study that included both PartoSure and fFN at a threshold of 50ng/ml (Nikolova 2015), both sensitivity and specificity (against the preterm birth within 7 days reference standard) were higher for PartoSure (80%, 95% CI 63.1, 91.6 and 94.6%, 95% CI 90.1, 97.5 respectively) compared to fFN at 50ng/ml (50.0%, 95% CI 21.1, 78.9 and 72.2% 95% CI 58.4, 83.5).[62, 63]

As would be expected, in the two studies assessing quantitative fFN at a variety of thresholds (APOSTEL-1 and EUIFS), as the threshold of fFN increased (<10, <50, <200 or <500 ng/ml) sensitivity decreased and the specificity increased (Table 8).[45, 46, 64]

2.3.2 Relevant systematic reviews

The test accuracy data for fFN at 50ng/ml presented above (Table 8) are based only upon the studies included in the systematic review of PartoSure, Actim Partus and quantitative fFN (see section 2.2). Recent systematic reviews were sought in order to identify other available data on fFN at 50ng/ml (either the older qualitative test or the modern quantitative test) in the prediction of preterm birth.

2.3.2.1 Data from systematic reviews of fFN at a threshold of 50ng/ml

One notable systematic review by Sanchez-Ramos et al. (2009), included 32 studies that used either a qualitative fFN test or the quantitative test at a threshold of 50 ng/ml.[67] In this review, pooled sensitivity (against a reference standard of delivery within 7 days) was 76.1% (95% CI 69.1, 81.9) and pooled specificity was 81.9 (95% CI 78.9-84.5).[67] More recently, Boots et al. (2014) published a systematic review that included both studies assessing fFN at a threshold of 50ng/ml and studies assessing cervical length measurement.[68] For fFN at 50ng/ml sensitivity and specificity estimates from 38 studies (against a reference standard of

delivery within 7 days) were similar to those reported in the previous review by Sanchez-Ramos (2009): in the Boots et al. (2014) review, pooled sensitivity was 75% (95% CI 69, 80) and pooled specificity was 79% (95% CI 76-83).[68] These values are also similar to those reported in recent NICE guidance, where across 20 studies of 'low' to 'very low' quality, sensitivity (against a reference standard of delivery within 7 days) ranged from 56% (95% CI not reported) to 100% (95% CI not reported) and specificity from 61.9% (95% CI 59.6, 62.5) to 92% (95% CI not reported).[26]

These systematic review data are also similar to the data for fFN at a threshold of 50ng/ml from the current overview (see section 2.3.1.2 and Table 8) which also ranged widely; sensitivity (against the 7 day reference standard) ranged from 23.8% (95% CI 17.3, 31.4) to 91.3% (95% CI 82.0, 96.7) and specificity ranged from 62.2% (95% CI 57.3, 66.9) to 99.1% (95% CI 97.3, 99.8).

2.4 Summary

Diagnostic test accuracy data were sought in two ways:

- A systematic review evaluating the test accuracy of the following index tests: PartoSure, Actim Partus, quantitative fFN at thresholds other than 50ng/ml
- A non-systematic overview of the test accuracy evidence, based on studies from the systematic review and supplemented with data from recent systematic reviews, of tests used in current clinical practice: fFN at 50ng/ml (qualitative or quantitative tests)

2.4.1 Data derived from the systematic review of diagnostic test accuracy

2.4.1.1 Included studies

Twenty studies met the systematic review inclusion criteria:

- Two 'comparative' studies (i.e. studies assessing more than one index test in the same population). One of these (APOSTEL-1)[45, 46] included both Actim Partus and quantitative fFN and the other (Hadzi-Lega 2017)[47] included both Actim Partus and PartoSure
- Fourteen studies assessing only Actim Partus (Abo El-Ezz 2014, Altinkaya 2009, Azlin 2010, Brik 2010, Cooper 2012, Danti 2011, Eroglu 2007, Goyal 2016, Lembet 2002, Riboni 2011, Tanir 2009, Ting 2007, Tripathi 2016, Vishwekar 2017)[1, 48-60]
- Three studies assessing only PartoSure (Bolotskikh 2017, Nikolova 2015 and Werlen 2015)[44, 61-63]
- One study assessing only quantitative fFN (EUIFS)[64]

All 20 studies evaluated diagnostic test accuracy against a reference standard of preterm delivery within 7 days.[1, 44-64] For seven studies (six Actim Partus studies and one PartoSure study), test accuracy was also measured against a reference standard of preterm delivery within 48hrs.[44, 51, 54, 55, 58-60]

In the studies assessing two index tests in the same sample (APOSTEL-1 and Hadzi-Lega 2017), sample sizes were 350 and 57 respectively.[45-47] Sample sizes in the other studies ranged from 30 to 468 for Actim Partus,[59, 60] 41 to 203 for PartoSure,[44, 62, 63] and the only study evaluating quantitative fFN alone included 455 participants.[64]

In addition, seven ongoing trials were identified which may be relevant to this review question, including four trials conducted in the UK (two of which aim to recruit over 1,000 participants).

2.4.1.2 Heterogeneity between studies

There was substantial methodological, clinical and statistical heterogeneity between studies including:

- Prevalence rate of preterm birth: prevalence of preterm delivery within 7 days ranged from 1.7% (0.6, 3.7) to 73.3% (60.3, 83.9) and within 48 hours from 2.4% (0.1, 12.9) to 58.3% (44.9, 70.9).[1, 44, 54]
- Mode of delivery: four studies reported that women who had a non-spontaneous delivery within the time-frame of the reference standard were excluded from the test accuracy data,[45-47, 61, 64] three further studies mentioned iatrogenic delivery as a reason for exclusion, but it is unclear how many of these deliveries occurred within the timeframe of the reference standard,[51, 60, 62, 63] and three studies report the number of spontaneous/iatrogenic deliveries but include the data from these women.[1, 50, 57] For the remaining ten studies, the mode of delivery was not clearly reported.[44, 48, 49, 52-56, 58, 59]
- Gestational age: the majority of included studies used 24 weeks as the lower limit for gestational age at enrolment,[1, 44-46, 48-54, 56-58, 64] with the lower limit in the remaining six studies ranging from 20 to 28 weeks.[47, 55, 59-63] The upper limit for gestational age varied more between studies, ranging from 32.6 weeks to 37 weeks gestation.[52, 57, 59, 60] No studies reported test accuracy data stratified by gestational age.
- Symptoms defined as indicative of preterm labour: all included studies state that women presented with symptoms indicative of preterm labour, and all but one study provided further detail regarding these symptoms.[58] All other studies reported uterine contractions as a necessary indicator of preterm labour.[44-57, 59-64] However there was

variation in the rate of uterine contractions necessary for inclusion.[45, 46, 48, 52-57, 59, 64] Other symptoms of preterm labour varied between studies, covering: abdominal or back pain, pelvic pressure, vaginal bleeding and/or vaginal discharge.

- Multiple gestations: four studies included women with multifetal pregnancies. [1, 45, 46, 60, 64] In these studies, multifetal pregnancies ranged from 6% to 20% of study participants.[1, 45, 46]
- Risk status: only one study clearly reports the risk status of participants.[61]. Heterogeneity of studies with regards to the risk status of women is, therefore, unclear.
- Dilation threshold and cervical length: all but four studies included a dilation threshold for exclusion.[1, 54, 55, 60] Typically the threshold was > or ≥ 3cm, but two studies had a lower threshold (>2cm and ≥2cm). [49, 56] In two studies, all included women had a transvaginal cervical length measurement <=30mm.[45, 46, 52]
- Other more specific exclusion criteria also varied between studies (e.g. cervical cerclage, previous tocolytic treatment, recent sexual intercourse, vaginal bleeding and prior cervical exam).
- Participant characteristics also differed between studies. These differences included average maternal age, gestational age at presentation, and history of preterm delivery.

2.4.1.3 Administration of index tests

Studies generally followed manufacturer's guidance on how to administer index tests. Key differences in how the test was administered include:

- Two studies used frozen samples in their analysis.[1, 45, 46] It is unclear how long samples were stored before testing. This protocol is inconsistent with manufacturer guidance and clinical practice.
- One Actim Partus study included two failed tests (no visible lines) as positive test results.[57]
- Two Actim Partus studies collected samples from the posterior fornix rather than the external cervix os.[45, 46, 58]

2.4.1.4 Provision of treatment

It should be noted that providing treatment (tocolytics and/or corticosteroids) may impact upon the occurrence of the reference standard (i.e. whether preterm delivery occurs) and this would impact upon the test accuracy data.

Whether a woman received treatment for symptoms of preterm labour varied substantially between studies. Moreover, the number of women receiving treatment was not always reported, particularly with reference to the results of the index tests.

This means that, in the included studies, it is difficult to ascertain the extent to which treatment may have impacted upon the test accuracy results.

2.4.1.5 Quality Appraisal

Phase 3 of the QUADAS-2 tool was used to evaluate the risk of bias and highlight concerns regarding applicability. All studies were single-gate DTA studies, and issues regarding risk of bias and concerns regarding applicability were minimal. However, the following key points were noted:

- Overall, there was a lack of detail regarding recruitment methods, with only five studies providing clear details.[50-52, 60, 62, 63]
- Two studies used frozen samples. Therefore, in these studies, the timing of the index tests was inconsistent with clinical practice and assessors could have potentially been aware of the reference standard (occurrence of pre-term birth within 48hrs or within 7 days).[1, 45, 46] There is also no clear evidence regarding the likely impact on the tests of longer-term storage.
- In eight studies there was lack of clarity regarding whether index test assessors were blinded to the results of additional diagnostic tests (e.g. cervical length).[45, 46, 52, 53, 56, 58, 59, 62-64] However, due to the nature of the index tests, there is little scope for bias to occur in their interpretation.
- In two Actim Partus studies, samples were collected from posterior fornix rather than external cervical os.[45, 46, 58]
- The lack of clarity regarding administration of tocolytics, particularly in reference to test results, precluded a thorough evaluation of the effect of treatment on test accuracy data.

2.4.2 Data derived from the overview of tests recommended in current NICE guidance

Several studies in the systematic review also provided data for tests currently used in clinical practice, i.e. fFN at 50ng/ml based on either a quantitative or qualitative test, (but these tests were not included in the systematic review of diagnostic test accuracy):

- Eight studies report diagnostic test accuracy data for fFN measured at 50ng/ml threshold (in addition to data for at least one index test).[1, 45, 46, 53, 56, 58, 59, 62-64]
- One of the studies reporting 50ng/ml fFN data, only does so for 66 out of the 203 patients recruited and included in the PartoSure analyses.[62, 63]
- In addition, recent NICE guidance, a systematic review of the test accuracy of fFN at a threshold of 50 ng/ml and a systematic review of the test accuracy of both cervical length measurement and fFN at a threshold of 50 ng/ml were included in the overview of these tests.[26, 67, 68] This was primarily to ensure that the data presented from primary studies was similar to that reported in recent systematic reviews.

2.4.3 Summary of the data available across the systematic review and overview

Table 9 summarises the results from the systematic review of diagnostic test accuracy for the three index tests (PartoSure, Actim Partus and quantitative fFN), and also summarises these results in the context of the test accuracy of fFN at 50ng/ml (data taken from the same studies).

As can be seen from Table 9, Actim Partus and PartoSure were assessed in the same sample in one study (Hadzi-Lega, 2017) and Actim Partus and quantitative fFN were assessed in the same sample in one other study (APOSTEL-1).[45-47] However, no studies were identified that assessed PartoSure and quantitative fFN in the same sample. Actim Partus was also assessed in the same sample as quantitative fFN at 50ng/ml in APOSTEL-1,[45, 46] in the same sample as the QuikCheck test in two studies (Tripathi 2016, Eroglu 2007),[53, 59] and in the same sample as an ELISA test in Riboni (2011).[56]

As well as being assessed in the same sample as Actim Partus in Hadzi-Lega (2017),[47] PartoSure was assessed in the same sample as the QuikCheck test in one study (Nikolova 2015, Table 9).[47, 62, 63] For quantitative fFN, data were also available at the 50ng/ml threshold in two studies (APOSTEL-1 and EUIFS, Table 9).[45, 46, 64]

Table 9 Summary of evidence and relative accuracy against the 7 day reference standard

		Actim Partus	PartoSure	Quantitative fFN		
				@10ng/ml	@200ng/ml	@500ng/ml
Index tests						
Actim Partus						
PartoSure		No difference (Hadzi-Lega 2017)				
Quantitative fFN	**@10ng/ml**	Sensitivity of fFN superior, specificity of Actim Partus superior (APOSTEL-1)	Indirect evidence only			
	@200ng/ml	No difference (APOSTEL-1)	Indirect evidence only			
	@500ng/ml	Sensitivity of Actim Partus superior, specificity of fFN superior (APOSTEL-1)	Indirect evidence only			
fFN at 50ng/ml						
Quantitative fFN at 50ng/ml		Specificity of Actim Partus superior, no difference in sensitivity (APOSTEL-1)	Indirect evidence only	Sensitivity of fFN @10ng/ml superior, specificity of fFN @50ng/ml superior (APOSTEL-1 and EUIFS)	Sensitivity of fFN @50ng/ml superior, specificity of fFN @200ng/ml superior (APOSTEl-1 and EUIFS)	Sensitivity of fFN @50ng/ml superior, specificity of fFN @500ng/ml superior (APOSTEl-1 and EUIFS)
QuikCheck		Sensitivity of Actim Partus superior & specificity of fFN superior (Tripathi 2016). However Eroglu 2007 showed no difference between tests.	Specificity of PartoSure superior, no difference in sensitivity (Nikolova 2015-note missing participants)	Indirect evidence only	Indirect evidence only	Indirect evidence only
ELISA		No difference (Riboni)	No evidence	Indirect evidence only	Indirect evidence only	Indirect evidence only

2.4.4 Summary of test accuracy data across the systematic review and overview

Table 10 summarises the sensitivity and specificity data for the index tests in the systematic review of test accuracy (PartoSure, Actim Partus, and quantitative fFN at thresholds other than 50ng/ml) as well as from the overview of tests recommended in current NICE guidance (qualitative fFN or quantitative fFN at 50ng/ml).[26]

With regard to the reference standard of preterm delivery within 7 days:

- Data set 1 (Table 10) reports test accuracy data obtained from the two studies that assess two index tests (included in the systematic review).
- Data set 2 (Table 10) reports the sensitivity and specificity values from all included studies (those evaluating only one index test and those evaluating more than one index test); for all index tests, data set 2 (Table 10) reports the range of sensitivities and specificities across the individual studies. Meta-analyses were conducted to calculate pooled sensitivity and specificity for the 16 studies assessing Actim Partus and for the four studies assessing PartoSure. Meta-analyses were not conducted for quantitative fFN because only two studies of this test were included.
- Data set 3 (Table 10) reports the range of sensitivities and specificities from included studies regarding the non-index tests that are currently recommended by NICE guidance (fFN at 50ng/ml).[26] For fFN at 50ng/ml, meta-anlyses were not conducted due to the heterogeneity across studies with regards to the tests used (QuikCheck fFN, ELISA, quantitative fFN).
- Data set 4 (Table 10) reports data from two relevant systematic reviews regarding a non-index test that is currently recommended by NICE guidance (fFN at 50ng/ml). Data from NICE guidance are also presented.[26]

With regard to the reference standard of preterm delivery within 48 hours:

- Data set 5 (Table 10) reports the range of sensitivities and specificities across the six Actim Partus studies included in the systematic review and also reports the sensitivity and specificity derived from the one PartoSure study providing test accuracy data against the 48hr reference standard. For Actim Partus, meta-analyses were conducted.

Table 10 Summary Table

Test Accuracy for the Prediction of Preterm Delivery within 7 days

Studies assessing more than one index test

Index Test	Source	Sensitivity % (95% CI)	Specificity % (95% CI)
fFN at 10ng/ml	Bruijn APOSTEL-1 (n=350)	95.7 (87.8, 99.1)	42.3 (36.5, 48.4)
fFN at 200ng/ml	Bruijn APOSTEL-1 (n=350)	71.0 (58.8, 81.3)	83.6 (78.8, 87.8)
fFN at 500ng/ml	Bruijn APOSTEL-1 (n=350)	42.0 (30.2, 54.5)	95.7 (92.7, 97.8)
Actim Partus	Bruijn APOSTEL-1 (n=350)	78.3 (66.7, 87.3)	89.3 (85.1, 92.7)
PartoSure	Hadzi-Lega 2017 (n=57)	83.3 (35.9, 99.6)	90.2 (78.6, 96.7)
Actim Partus	Hadzi-Lega 2017 (n=57)	83.3 (35.9, 99.6)	76.5 (62.5, 87.2)

Studies assessing a single index test

Index Test	Source	Sensitivity % (95% CI)	Specificity % (95% CI)
Actim Partus	Pooled (16 studies)	77 (68, 83)	81 (76, 85)
	Range (16 studies)	33.3 (4.3, 77.7) - 94.7 (89.9, 97.7)	50.0 (24.7, 75.3) - 93.5 (82.1, 98.6)
PartoSure	Pooled (4 studies)	83 (61, 94)	95 (89, 98)
	Range (4 studies)	0 (0.0, 97.5) - 100.0 (73.5, 100.0)	90.2(78.6, 96.7) - 97.5(96.8, 99.9)
fFN at 10ng/ml	Range (2 studies)	93.8 (82.8, 98.7) - 95.7 (87.8, 99.1)	32.2 (27.7, 37.0) - 42.3 (36.5, 48.4)
fFN at 200ng/ml	Range (2 studies)	70.8 (55.9, 83.0) - 71.0 (58.8, 81.3)	78.6 (74.3, 82.5) - 83.6 (78.8, 87.8)
fFN at 500ng/ml	Range (2 studies)	29.2 (17.0, 44.1) - 42.0 (30.2, 54.5)	94.3 (91.6, 96.4) - 95.7 (92.7, 97.8)

Supplementary data from included studies

Test	Source	Sensitivity % (95% CI)	Specificity % (95% CI)
fFN at 50ng/ml	Range (8 studies)	23.8 (17.3, 31.4) - 91.3 (82.0, 96.7)	62.2 (57.3,66.9) - 99.1 (97.3,99.8)

Data extracted from Systematic Reviews

Test	Source	Sensitivity % (95% CI)	Specificity % (95% CI)
fFN at 50ng/ml	Sanchez-Ramos 2009 Pooled (32 studies)	76.1 (69.1,81.9)	81.9 (78.9, 84.5)
fFN at 50ng/ml	Boots 2014 Pooled (38 studies)	75 (69, 80)	79 (76, 83)
fFN at 50ng/ml	NICE 2015 guidance Range (20 studies)	56[a] -100[a]	61.9 (59.6, 62.5) - 92[a]
Test Accuracy for the Prediction of Preterm Delivery within 48 hours			
5. Studies assessing a single index test			
Index Test	**Source**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**
Actim Partus	Pooled (6 studies)	87 (74, 96)	73 (62, 82)
	Range (6 studies)	65.7 (47.8, 80.9) – 100 (47.8, 100.0)	56.0 (34.9,75.6) - 82.4 (56.6, 96.2)
PartoSure	Werlen 2015 (n=41)	0.0 (0.0, 97.5)	97.5 (86.8, 99.9)

Key: [a] 95% CI not reported

3 Assessment of clinical effectiveness (end-to-end) studies

End-to-end studies investigate the clinical impact of performing tests by following patients from testing, through treatment, to final clinical outcomes. RCTs provide the best quality end-to-end comparative evidence, providing a direct link between a testing strategy and the clinical outcomes of interest. We performed a systematic review of end-to-end studies, with a particular focus on RCTs but also including other controlled study designs.

3.1 Methods for reviewing effectiveness

3.1.1 Identification of studies

The same searches were performed as for the review of diagnostic accuracy studies (section 2.1.1.1). In brief, these included searches of: electronic databases (these were designed to identify all studies assessing PartoSure, Actim Partus and quantitative fFN); all systematic reviews identified by the electronic searches; trial registries; Google Advanced; reference lists of included DTA studies; studies citing the included DTA studies; and industry submissions to NICE.

As with the review of diagnostic test accuracy (section 2.1.1), screening for relevant studies was in two stages (screening of titles and abstracts and then screening of papers obtained in full). At both stages, screening was performed concurrently with the screening for the review of test accuracy studies and this was done independently by two reviewers (two of JVC, SD, MB, and HC). Pre-specified inclusion and exclusion criteria were used (see Section 3.1.2 below). Disagreements were resolved by discussion.

3.1.2 Inclusion and exclusion criteria

3.1.2.1 Population

With regards to the population, inclusion criteria were the same as that for the review of test accuracy studies (see section 2.1.1.2.1).

3.1.2.2 Interventions

The interventions under consideration were identical to those in the review of test accuracy (i.e. PartoSure, Actim Partus and quantitative fFN; see section 2.1.1.2.2).

3.1.2.3 Comparators

Studies were eligible for inclusion if at least one of the interventions was compared with one or more of the following comparators:

- One of the other interventions (with or without an assessment of clinical symptoms)
- The quantitative fFN test used with a threshold of 50 ng/ml (with or without an assessment of clinical symptoms)
- A qualitative fFN test (with or without an assessment of clinical symptoms)
- Clinical assessment of symptoms alone

3.1.2.4 Outcomes

In accordance with the NICE scope,[13] eligible studies should have included one or more of the following outcomes in order to be eligible for inclusion:

- Perinatal mortality
- Neonatal morbidity and mortality
- Long-term health problems in the child
- Maternal morbidity and mortality
- Health-related quality of life
- Anxiety associated with confidence in the test results
- Number of women admitted to hospital
- Number of re-presentations to hospital within 48 hours and 7 days
- Number of women who have tocolytics / corticosteroids
- Length of inpatient hospital stay
- Number of transfers of pregnant women and neonates between hospitals
- Time to delivery from presentation
- Number of women treated with maternal corticosteroids appropriately (that is, they deliver within 7 days following treatment)
- Number of women treated with maternal corticosteroids inappropriately (that is, they do not deliver within 7 days following treatment)
- Impact on neonatal intensive care resource planning
- Gestational age at birth

Studies that report data on costs only were not eligible for inclusion in the review of clinical effectiveness.

3.1.2.5 Study design

RCTs were primarily sought for this review. However, other controlled designs (prospective or retrospective) were also eligible for inclusion.

3.1.3 Other methods

Further aspects of the review methods (data extraction strategy, critical appraisal strategy and methods of data synthesis) are not described as there were no included studies.

3.2 Results

3.2.1 Quantity and quality of research available

After screening 2,623 items, no studies were identified that met the inclusion criteria for the review of clinical effectiveness. This was because none of the studies compared the tests of interest with a comparator with regards the clinical outcomes of interest; there were no studies identified where some women received one test and some received another and even in the studies identified in the test accuracy review where women received more than one test (see section 2.3) there was no clear indication that treatment decisions were based on the results of one test for some women and based on the results of the other test for other women. Indeed, these studies did not provide data on the clinical outcomes of interest.

3.2.2 Assessment of effectiveness

We were not able to draw any conclusions on the effectiveness of PartoSure, Actim Partus or fFN from the systematic review of end-to-end studies. Our broad searches were not restricted by a study design filter and were focused on identifying all studies of the tests of interest, and this makes it unlikely that we have missed major items of published literature. In order to identify other (potentially unpublished) literature, and to reduce the likelihood of overlooking any relevant end-to-end studies, web searches and searches of trial registries were conducted. We also considered conference abstracts that were identified in the electronic searches, but from the limited information provided in these abstracts, it did not appear that any useful end-to-end data were available.

It is, however, important to consider that this review was looking for evidence from controlled study designs. For the systematic review of clinical effectiveness, we did not look for evidence of clinical effectiveness from other designs (e.g. uncontrolled pre-post studies). This decision was made because these designs may be too open to bias to be worth including in a systematic review of end-to-end studies, even if they are the only evidence available. It should be noted that it may be necessary for data from pre-post studies to be used in economic modelling (i.e. to parameterise a model when this is the only evidence

available) and these data may, therefore, be obtained from studies that have not been selected via a systematic reviewing process.

3.3 Summary

We were not able to draw any conclusions on the effectiveness of PartoSure, Actim Partus or fFN from the systematic review of end-to-end studies.

It is unlikely that we have missed major items of published literature; our broad searches were not restricted by a study design filter and were focused on identifying all studies on the tests of interest. In order to identify other (potentially unpublished) literature, and to reduce the likelihood of overlooking any relevant end-to-end studies, web searches and searches of trial registries were conducted. We also considered conference abstracts that were identified in the electronic searches, but from the limited information provided in these abstracts, it did not appear that any useful end-to-end data were available.

It is, however, important to consider that this review was looking for evidence from controlled study designs and did not look for evidence from other designs (e.g. uncontrolled pre-post studies). It is, therefore, worth considering whether conducting controlled studies in this area can reasonably be expected. On balance, it does not seem unreasonable to expect such studies in this population; the principal barrier to conducting an RCT would be the potential difficulty of recruiting participants during an acute medical situation (e.g. time needed to consent and randomise). However, this population and these tests would also lend themselves well to an RCT design with regards the length of follow-up required (for a number of key outcomes the length of follow-up could be less than a year).

Nevertheless, the decision to only include controlled studies was based primarily on the fact that uncontrolled designs may be too open to bias to be worth including in a systematic review of end-to-end studies, even if they are the only evidence available. It should be noted, however, that it may be necessary for data from pre-post studies to be used in economic modelling (i.e. to parameterise a model when this is the only evidence available) and these data may, therefore, be obtained from studies that have not been selected via a systematic reviewing process.

4 Data informing the economic modelling

As described above, the systematic review produced limited diagnostic test accuracy data (Chapter 2) and no clinical effectiveness data (Chapter 3) for populating an economic evaluation of diagnostic tests of interest. There was no single diagnostic test accuracy study that evaluated all three index tests, and only two studies (APOSTEL-1 and Hadzi-Lega et al. 2017)[45-47] compared at least two index tests. There is a high degree of heterogeneity between the reviewed diagnostic accuracy studies in terms of prevalence of preterm birth, mode of delivery, gestational age, definition (symptoms) of preterm labour (including dilation threshold), multiple gestations, participant characteristics and provision of treatments. In the light of this, comparisons among tests on the basis of the results of the meta-analyses presented in this chapter are likely to be biased, since the studies providing data for meta-analyses are very different both within and between the different tests. Therefore, of the studies identified and reviewed in chapter 2, only studies that presented results for at least two different index tests in the same patient sample were used for the economic evaluation in Chapter 6. There were two such studies, APOSTEL-1 which assessed both Actim Partus and fFN and Hadzi-Lega (2017) which assessed Actim Partus and PartoSure.[45-47] In addition, we further excluded studies that investigated laboratory-based ELISA qualitative (at 50 ng/ml) fFN tests, as this technology is no longer in use; meta-analysis of the remaining 4 studies of Actim Partus vs qualitative fFN[45, 46, 53, 58, 59] thus provided the diagnostic test accuracy results used in the economic evaluation. Details are presented in section 6.

5 Systematic review of existing cost-effectiveness evidence

The first part of this chapter presents the results of a systematic review of previous economic studies on the diagnostic test interventions. In the second part we also present a review the methods used in previous evaluations of diagnostic tests for diagnosis of preterm labour in symptomatic women with intact membranes.

5.1 Methods for reviewing economic evaluation studies

Systematic review methods were used to identify previously published economic evaluations of the three tests under consideration; PartoSure, Actim Partus, and quantitative fFN. The review was undertaken following the general principles published by the University of York Centre for Reviews and Dissemination (CRD).[40]

5.1.1 Identification of studies

Methods followed those reported in section 2.1.1 for study identification except only electronic databases searches were conducted. Studies were screened by two reviewers (RMM and JVC).

5.1.2 Inclusion and exclusion criteria

Population, index test and reference standard matched those reported in section 2.1.2. However, for the economic evaluations review, the criteria for inclusion was studies that reported healthcare costs of an index test without restriction in terms of study design.

5.1.3 Data extraction strategy

Data were extracted by one reviewer (RMM) using standardised data extraction templates.

5.1.4 Critical appraisal strategy

The quality of the studies was assessed in detail by an experienced health economist (RMM) according to the criteria specified by the CHEERS checklist.[69]

5.1.5 Methods of data synthesis

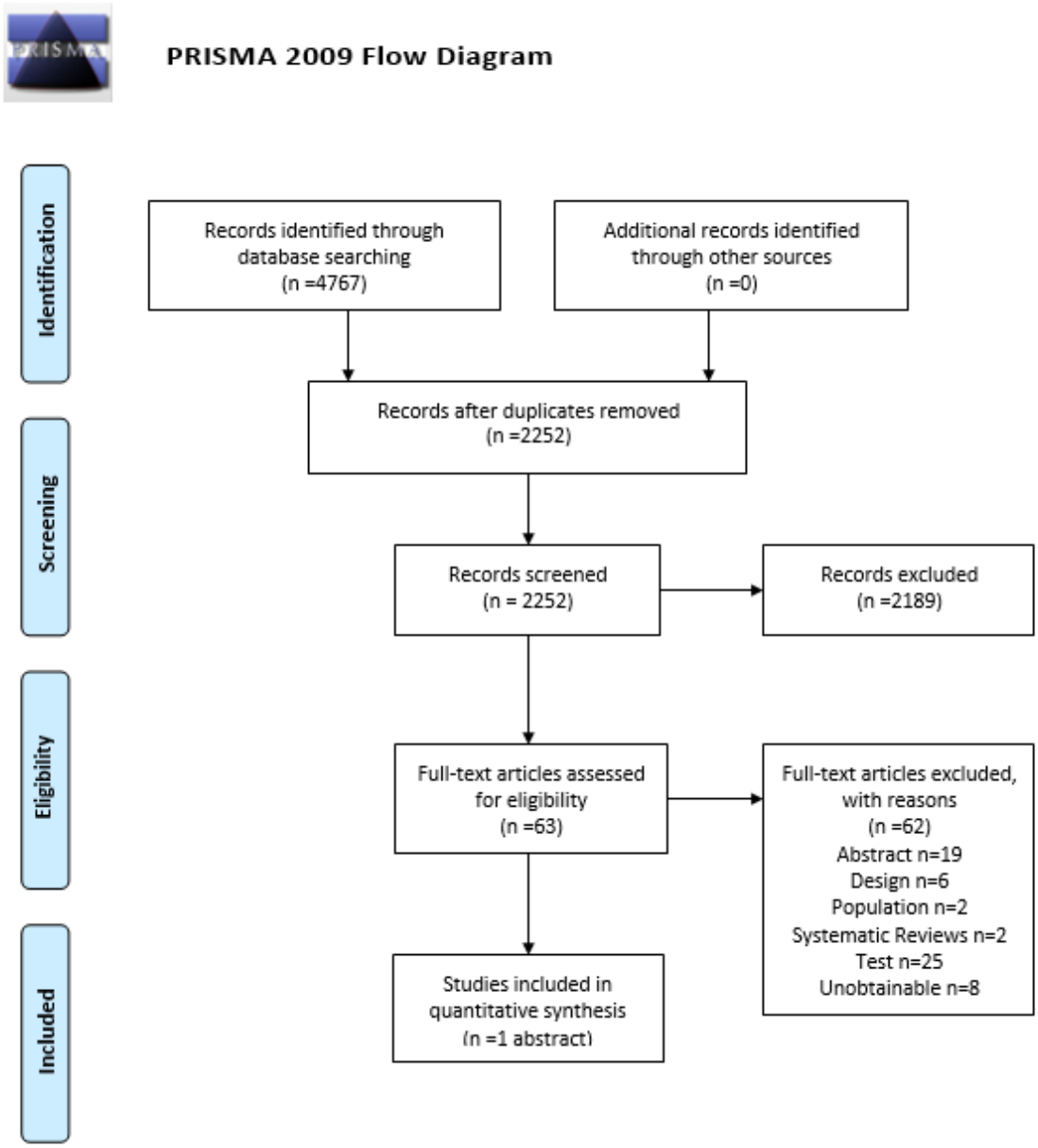
Data were narratively reported. Methods and results were tabulated in the prices and currencies as reported by the identified studies.

5.2 Results

Figure 7 shows the study flow diagram of this review. The electronic database search identified 2252 records after deduplication. All were screened on title and abstract, of these 63 citations were taken to full-text screening. One study met the inclusion criteria. It was a

conference abstract of an unpublished MSc dissertation.[70] We contacted the authors and they provided a copy of the dissertation, which is the basis for this review.

Figure 7 PRISMA flow diagram for economic evaluation review[71]



5.3 Economic evaluation studies

The only study included in our review reported the management practice with the qualitative fFN test at the preterm clinic of St. Thomas' Hospital in London. It also evaluated the hypothetical use of antenatal corticosteroids (ACS) and tocolysis with full compliance with the treatment protocol at different fFN thresholds (positive result; ≥10ng/ml, ≥50ng/ml, ≥200ng/ml and ≥500ng/ml) provided by the Hologic Rapid 10Q System against delivery

outcomes. Clinicians were blinded to the quantitative fFN concentration “to prevent it influencing their management based on the qualitative fFN test result” (Table 11).[70]

The study reported the proportion of compliance with the qualitative fFN treatment protocol: 67% of positive cases (35/51) were treated with ACS whilst 6% of negative cases (16/252) were given treatment. Two (6%) and 10 (29%) out of the 35 women who had a positive test result and were treated with ACS delivered within 7 days and before 37 weeks, respectively.

In addition the study analysed the rate of compliance with the protocol of administering tocolysis treatment to women with a positive test result. Only 14% (10/75) women testing positive with qualitative fFN were administered tocolytics, while 2% (6/282) of women testing negative received tocolytics. Of those patients who tested positive and were given tocolytics, one out of 10 delivered within 7 days, whilst four out of 10 delivered before 37 weeks.

In the published abstract, results are presented for the number needed to prevent one case of RDS for the ‘no test and treat all’ option and the 200 ng/ml fFN threshold option, respectively equal to 1540 and 80. However the methods used to obtain these numbers are not given in the abstract nor any reference to these results appears in the dissertation. The dissertation does provide however detailed information on some of the data required to calculate those numbers for the different diagnostic and treatment options, in the form of numbers needed to successfully administer steroids to 1 woman delivering within 7 days of testing (no test and treat all, 77; 200 ng/ml, 9) and women delivered before receiving a full steroids course (i.e. within 24 hours of testing; 3 in both cases).

5.3.1 Critique

Although the study by Gibson and colleagues did not aim to assess the cost-effectiveness of the different diagnostic strategies, it did provide information with which to model the cost-effectiveness of the following two sets of comparisons: a) no test and treat all with steroids vs. qualitative testing with fFN and treat those with positive results, and b) testing options investigated at the quantitative fFN thresholds 10, 50, 200 and 500 ng/ml.[70]

Given the available data from the study, the costs per patient adequately treated with steroids (i.e. within 7 days) were calculated by the AG using the following formula:

For the comparison between no test and treat all with steroids:

Incremental cost per additional patient adequately treated

$$= \{C_{ACS} - \Big(C_{fFN} + fFN_{+} * (C_{ACS} + C_{H}) + fFN_{F-} * (C_{ACS} + C_{H})\Big)\} \frac{N}{N_{TA} - N_{fFN}}$$

$$= \{C_{ACS} - \Big(C_{fFN} + (fFN_{+} + fFN_{F-}) * (C_{ACS} + C_{H})\Big)\} * NNT$$

Where:

- C_{ACS} : cost of steroids,
- C_{fFN} : cost of fetal fibronectin test,
- fFN_{+} : probability of a positive test result,
- C_{H} : costs of hospital admission
- N : total sample size
- N_{TA} : number of mothers given ACS within 7 days & at least 1 day before delivery in the treat all strategy
- N_{fFN} : number of mothers given ACS within 7 days & at least 1 day before delivery in the testing strategy
- fFN_{F}- : probability of a false negative test result.
- NNT : number needed to treat to avoid one case of inadequate treatment without testing

Thus the incremental cost per patient adequately treated is equal to the number needed to treat to avoid one case of inadequate treatment without testing (NNT) times the incremental cost per patient of treating all versus testing and treating positive cases. By combining the data from Gibson and colleagues[70] with treatment effectiveness data and cost and utility values used by previous models, e.g. NICE 2015 below,[26] we can obtain the incremental cost per case of RDS avoided, cost per life saved, and, subject to the natural reservations about projecting long term outcomes, cost per QALY gained. The formulae become:

Incremental cost per RDS case avoided =

$$= \{C_{ACS} + C_H - \left(C_{fFN} + (fFN_+ + fFN_{F-}) * (C_{ACS} + C_H)\right)\} * \frac{NNT}{ARR_{RDS}}$$

Incremental cost per IVH case avoided =

$$= \{C_{ACS} + C_H - \left(C_{fFN} + (fFN_+ + fFN_{F-}) * (C_{ACS} + C_H)\right)\} * \frac{NNT}{ARR_{IVH}}$$

Incremental cost per death avoided =

$$= \{C_{ACS} + C_H - \left(C_{fFN} + (fFN_+ + fFN_{F-}) * (C_{ACS} + C_H)\right)\} * \frac{NNT}{ARR_{Death}}$$

Incremental cost per QALY gained =

$$= \left\{C_{ACS} + C_H - \left(C_{fFN} + (fFN_+ + fFN_{F-})(C_{ACS} + C_H)\right)\right\} * \frac{NNT}{ARR_{Death} MaxQALY + ARR_{RDS} D_{RDS} Disu_{RDS} + ARR_{IVH} D_{IVH} Disu_{IVH}}$$

Where:

- ARR_{RDS} is the absolute risk reduction (i.e. the difference between the absolute probability with and without appropriate ACS treatment administration) of RDS occurring.
- ARR_{IVH} is the absolute risk reduction of IVH occurring after steroid treatment.
- ARR_{Death} is the absolute overall death risk reduction (which includes the reduction in death mediated through RDS and IVH).
- D_{IVH} is the conditional probability of death in neonates with IVH.
- D_{RDS} is the conditional probability of death in neonates with RDS.
- $Disu_{IVH}$ is the QALY loss from IVH
- $Disu_{RDS}$ is the QALY loss from RDS

Adopting the values in the NICE Guideline model discussed below and summarised in Table 12 one may calculate the relevant incremental cost-effectiveness measures using the diagnostic test results reported by Gibson and colleagues.[70]

Table 11 Included cost effectiveness study

Study & period of study	Population	Country & setting	Test/ diagnostic strategy	Study design	N	Time-frame	Outcome and accuracy test results (if available)	Results	Comments
Gibson, Shennan and Hazelgrave 2013 (un-published)[70]	Subsample of EQUIPP: High-risk symptomatic women age 18 years and older between 23^{+0} and 34^{+6} weeks gestation, (GA range suitable for ACS and tocolysis administration). Presenting symptom: TPTL (contractions >2 in 30 min) 15.2%; abdominal pain 70.9% of women; remaining 13.9% presented with other symptoms such as tightening and pelvic pressure.	Pre-term Surveillance Clinic at St Thomas' Hospital in London.	A. Management without fFN Testing (treat all) B. qualitative fFN protocol in reality (<full compliance) C. full adherence to qualitative fFN Protocol D. Modelled implementation of quantitative fFN at 10, 50, 200 and 500 ng/ml thresholds Clinicians blinded to quantitative fFN results. Aim was to evaluate whether quantitative fFN could add value to clinical management protocol	Prospective cohort	ACS analyses: 306 Tocolyses analyses: 351	Until delivery	sPTB <34 weeks gestation and <37 weeks gestation; delivery within 7 days and 14 days of testing Appropriate management: the number of symptomatic women given the intervention, NNT, and number of cases missed (spontaneous delivery within the specified time-frame and did not receive intervention)	Additional NNT to successfully administer steroids to one woman within 7 days of testing Figures not calculated by the authors	The analysis of ACS also included women who delivered within 24 hours of fFN testing as missed cases, regardless of whether they received ACS, because of the evidence "suggesting ACS is ineffective in the reduction of RDS when delivery occurs within 24 hours of treatment (RCOG, 2010b)." On the finding that the 500 ng/ml threshold results in 12 fewer women treated with ACS and 1 missed case in need of treatment relative to the 200 ng/ml, whereas no missed cases occur at the 200 vs lower thresholds, the authors concluded that the optimal risk: benefit threshold for ACS use is 200 ng/ml. The corresponding analysis for tocolysis led them to choose the 500 ng/ml threshold to indicate its use.

Notes: * At least one of the following risk factors: History of previous PTB (<37 weeks gestation)/ second trimester loss (≥16 weeks gestation); Short cervical length (<25mm) measured on ultrasound at 18+0 – 27+6 weeks gestation; Previous cervical surgery; Exclusion Criteria: Congenital abnormality; Sexual intercourse within 24 hours; Blood-stained swap; previously administered steroids/tocolysis; Symptomatic visit number ≥2;

Key: ACS, antenatal corticosteroids; fFN, fetal fibronectin; GA , gestational age; NNT, Number needed to treat; RCOG, Royal College of Obstetricians and Gynaecologists; RDS, respiratory distress syndrome; sPTB, spontaneous preterm birth; TPTL, threatened preterm labour

Table 12 Model parameter values from the NICE Guideline model (NICE 2015)[26]

Parameter	Parameter definition	Values	Source
C_{ACS}	Cost of full antenatal corticosteroid course	Included in Cost of hospital (C_H)	1
C_{fFN}	Cost of fFN test	£37.50	1
C_H	Cost of hospital admission	£1,050	1
fFN_+	Marginal probability of positive fFN test result	0.18	2
fFN_{F-}	Marginal probability of false negative fFN test result	0.0082	2
NNT	Number needed to treat to avoid one inadequately treated case	28.5	2
ARR_{Death}	Absolute risk reduction of death from treatment	0.049	1
ARR_{RDS}	Absolute risk reduction of RDS from treatment	0.052	1
ARR_{IVH}	Absolute risk reduction of IVH from treatment	0.015	1
D_{RDS}	Death risk from RDS	0.054	1
D_{IVH}	Death risk from IVH	0.300	1
MaxQALY	Maximum lifetime QALYs without RDS or IVH	22.44	1
$Disu_{RDS}$	QALY loss from RDS	3.85	1
$Disu_{IVH}$	QALY loss from IVH	4.5	1

Notes: [1] NICE Guideline 2015[26] [2] Gibson et al. 2014 [70]
Key: IVH, intraventricular haemorrhage; fFN, fetal fibronectin; NICE, National Institute for Health and Clinical Excellence; QALY, quality-adjusted life year; RDS, respiratory distress syndrome.

These values result in an incremental cost per case of IVH avoided of £1,548,291, incremental cost per RDS avoided of £466,622, incremental cost per death avoided of £473,967 and incremental cost per QALY gained of £20,942 with the 'No test and treat all' strategy relative to fFN. These figures do not account of any negative effects of inappropriate use of steroids on the infant's health, and may therefore be considered a lower bound estimate.

Gibson and colleagues found that variation of the threshold from 10 to 50 to 200 ng/ml resulted in the same number (n=2) of false negative cases that delivered within 7 days of testing.[70] In terms of cost effectiveness analysis, that finding means that qualitative testing using the 10 and 50 ng/ml threshold are dominated by testing at the 200 ng/ml threshold since the latter results in lower resource use for testing and hospital admissions than the lower threshold strategies. Moving from the 200 to the 500 threshold, however, resulted in one additional missed preterm birth case.[70] Using the corresponding formula for comparing two strategies using successively increasing thresholds of the quantitative fFN test, the assessment group (AG) calculates that the incremental costs per RDS, IVH, and death avoided of using the lower threshold is, respectively, £221,115, £770,000, and £235,714. The incremental cost per QALY is £10,415. Therefore at the £20,000 NICE cost-effectiveness threshold, the optimal, cost-effective diagnostic strategy is to use the quantitative fFN with a threshold of 200 ng/ml.

It is evident from the small numbers of false negative cases just presented that the findings from the study by Gibson and colleagues are highly uncertain.[70] This also highlights the need for evidence synthesis over multiple studies in order to derive meaningful evidence.

We highlight that the above formulae allow for a separate treatment of the costs of hospital admission and steroid treatment, in contrast with other models discussed below. This may be important since diagnostic guidelines or protocols being used in some centres, e.g. Guy's and St Thomas' Hospital (London), suggest that the fFN threshold concentrations used by clinicians in the obstetrics department to decide when to admit a patient may be different from those used by them for deciding when to administer steroids. Therefore the cost of treatment (i.e. the sum of steroid costs C_{ACS} and hospital costs C_H) may vary across different quantitative fFN thresholds.

5.3.2 Summary

One abstract was identified that investigated some measure of costs or cost-effectiveness of the interventions of interest to this assessment. In this section, we have reviewed the abstract and corresponding dissertation that reported results in terms of the number needed to treat to achieve a desired neonatal outcome, and we have shown how these data may be used in conjunction with the literature to derive useful information about the cost-effectiveness of different thresholds for the quantitative fFN test. We note that current treatment protocols in some hospitals may allow for the use of different quantitative fFN thresholds to decide whether to administer steroids and admit to hospital.

5.4 Observational cost minimisation studies

A set of studies was found in the systematic search of electronic bibliographic databases that investigated the healthcare costs of the comparator in this review, qualitative fFN testing vs. no test and treat all. These studies tended to date from 8 to 10 or more years ago and include implementation evaluations. Although they are not relevant to our main study question, i.e. the evidence on index tests, these studies provide some background evidence on the role of operational factors in the costs and cost-effectiveness of interventions in routine practice.

The study by Abenheim and colleagues describes the cost implications for the addition of rapid fFN to clinical examination in a tertiary university hospital in Montreal, Canada.[72] The diagnostic protocol was clinical evaluation of women presenting with symptoms of threatened preterm labour followed by fFN testing in pregnant women without a confirmed (i.e. cervix dilated >3cm in presence of contraction) or ruled out (cervix was closed and uneffaced and monitoring revealed no palpable or measured contractions) preterm labour diagnosis by clinical examination versus clinical examination alone.

This study evaluated fFN in a setting where cervical length measurement has not been incorporated as part of the assessment of patients presenting to a labour and delivery unit with preterm contractions. The fFN group was a prospective cohort of 116 pregnant women, of whom 36 were tested for fFN. Three patients had a positive test result, of whom one delivered within 7 days after admission (33% PPV) and the other two were eventually discharged from hospital without delivery. Thirty-three pregnant women had a negative test result, none of whom delivered within two weeks (100% NPV); three of these women (9%) were however admitted to hospital. The latter is one of the major strengths of this study, given the paucity of evidence on the effect of testing on patient management in this area.

The authors acknowledged the absence of pharmaceutical and radiological and laboratory costs as a limitation of their study.[72] In this regard, the study does not provide evidence on the proportion of women who were managed adequately with corticosteroid treatment (i.e. within 7 days of delivery) or indeed on the overall proportion of preterm deliveries including those beyond two weeks after testing. The authors also state that fFN was overused during the study period, thus preventing an accurate assessment of the proportion of women that would have required additional evaluation if testing had been unavailable.

A US study compared the number of hospital admissions in the year after the adoption of a laboratory based fFN protocol with the baseline 12 month period in a single provider and its tertiary referral centre, covering the period July 1995-June 1997.[73] The protocol specified that those with a negative test result should be asked to return 2 weeks later for re-examination and testing. Adopting fFN reduced the number of admissions from 28.1% to 17%, the preterm labour admissions per patient from 1.8 to 1.6 (p=0.002), and the proportion of patients with tocolytic therapy from 10% to 7.9% (p=0.030). This study also reported neonatal outcomes, i.e. % neonatal intensive care unit (NICU) admissions, median days of NICU length of stay, ventilation duration, % steroid administration among infants admitted to NICU, but these were not reported in a manner useful to our purposes. In any case, the fFN testing protocol in this study is outdated since it was based on a laboratory assay (as opposed to the rapid fFN test commonly used these days) and required that mothers testing negative return two weeks later for fFN re-testing.

5.4.1 Other studies with relevant outcomes

Bergella and colleagues'[74] systematically reviewed the RCT evidence on fFN and found that it resulted in an increasing trend towards admission to NICU (RR 2.48, 95% CI 0.96, 6.46) relative to clinical examination alone (blinded to fFN results), which had a 7.45% prevalence across the two RCTs reporting this outcome.[75, 76]

5.5 Model-based studies

We identified six different model structures presented in modelling studies of diagnostic interventions of pre-term labour: Deshpande et al.; Honest et al.; Chuck and Nguyen; Boyd et al.; Mozurkewich et al; the 2015 NICE guidelines for preterm labour; and van Baaren.[77-82] We describe these models as presented in their most recent applications found in the published literature (Table 13). These are all decision tree models, which vary in four principal aspects. The first is the type of study (cost minimisation, cost effectiveness, and cost-utility analysis); the only cost-utility model was that developed for the 2015 NICE Guideline by the Royal College of Obstetricians and Gynaecologists (RCOG).[26] The second aspect is the length of analytical horizon, where some models measured outcomes until delivery, thus assuming no differences beyond that landmark between diagnostic strategies whereas other models assessed outcomes until neonatal death or hospital discharge or in one case extrapolated neonatal outcomes to lifetime. The third aspect is the obstetrician's compliance, where the model assumes perfect compliance with the treatment protocol based on the diagnostic test results (i.e. all positive cases are treated and no negative cases are treated), as opposed to accounting for the behavioural factors that reduce compliance with those protocols. The fourth characteristic is the treatment being modelled. One model[26] assumes that all positive cases are treated with tocolytics, whereas other models base their modelling of neonatal health outcomes on the use of steroids independently of tocolytic usage.

5.5.1 Detailed review of individual models

5.5.1.1 Cost minimisation studies

A decision analysis was used by Chuck and Nguyen (2015) to evaluate the health system costs following the adoption of fFN testing in Alberta, Canada, in January 2008.[79] Their evaluation used observational data from inpatient and outpatient administrative medical records covering the period April 2002-March 2013. It linked data from the provincial laboratory system to determine the proportion of episodes presenting with signs and symptoms of preterm labour that resulted in: admissions, hospital transfers, preterm birth (<37 weeks) with false labour, and fFN testing.

The study analysed the proportion of transfers between those who received fFN and those who did not from a lower level unit to a tertiary care unit. The rate of admissions was also analysed using the outpatient administrative data. The inpatient data was used to analyse the length of stay.

The model divided the episodes of pregnant women presenting with signs and symptoms of preterm labour between true preterm labour episodes and episodes who did not deliver before 37 weeks (see Figure 8). For each of these subgroups the decision between performing the rapid fFN test and not performing the test was evaluated. The model was populated with parameter values from logistic regression including maternal and patient management characteristics covariates. The main model parameters for our purposes are summarised in Table 14.

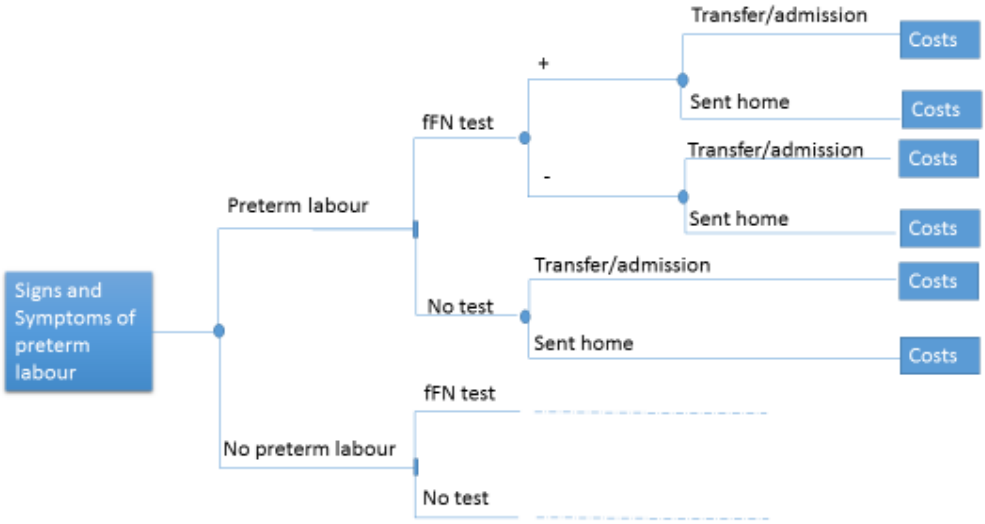
Table 13 Modelling studies

Study	Population	Perspective	Setting	Test/diagnostic strategy	Model structure	Time-frame	Effectiveness and Costs parameters	Type of study	Comments
Deshpande et al. 2013[77] *Update of Honest et al. 2009*[78]	Threatened PTL	NHS	Hospital	fFN & clinical examination vs clinical examination alone	Decision tree	Before delivery	Steroids and tocolytics costs Admission costs (Length of stay) Hospital transfers Ultrasound (after admission)	Cost minimisation	Documents the use of steroids and tocolytics from UK diagnostic study data reported by Dutta and Norman 2011 Useful source of data on cost parameters Does not include cost of delivery
Chuck & Nguyen 2015[79] *April 2002 – March 2013*	Threatened PTL or early onset of delivery in administrative databases (inpatient & outpatient) Alberta, Canada	Health system	Hospital	fFN vs no fFN	Decision tree	Delivery	Test specificity & sensitivity Testing rates Admission rates after testing Transfers LOS Health care costs	Cost minimi-sation	Populated with data from observational study (rates of admission and transfers). The observational study found that fFN increased transfers and admissions Assumed no costs savings from reductions of hospital admissions for episodes with false - results, as costs would only be delayed.
Boyd et al 2011[80]	Threatened PTL (clinical diagnosis) 24+ weeks UK	NHS	Hospital	fFN vs no test	Decision tree	Three months post-birth or neo-natal hospital discharge	Hospitalisations Transfers Healthcare costs Neonatal morbidity & mortality	Cost-effectiveness	fFN saved costs but had a “small but potentially detrimental” increase in neonatal morbidity, and a “negligible increase in mortality”. It did not measure negative effects of steroids use in false +. Only pre-term infants subject to morbidity are exposed to mortality risk.
Mozurkewich et al. 2000[81]	Threatened PTL (regular uterine contractions) 24 to 34 weeks, intact membranes, w/o	Third party payer	Tertiary care unit	Rapid fFN vs. treat all w/ steroids as outputs	Decision tree	Neonatal hospital discharge or death	Total cost = Triage or outpatient + fFN testing + hospitalization & treatment + Maternal delivery + neonatal care	Cost-effective-ness Cost per case of RDS/ neonatal	Maternal tocolytic side effects were not measured. Assumed that women having side effects necessitating discontinuation of one tocolytic would be given another tocolytic if necessary. Thus maternal side effects would

	cervical dilation ≥3 cm. US						RDS neonatal deaths	death avoided	not be related to the final probabilities of RDS and neonatal death. Assumed that infants with RDS receive surfactant. It did not add costs for maternal or neonatal transport.
NICE Diagnosis and treatment of Preterm labour and birth 2015[26]	Women with suspected PTL and intact membranes England	NHS	Hospital	Treat all vs.test & treat positive cases	Decision tree	Lifetime of new-born	QALY Costs (tests, treatment, maternal & neonatal admissions, lifelong healthcare), Sensitivity and specificity of SPTB within 7-days Mortality RDS IVH	Cost-utility	What if analysis accounted for differences in costs and benefits by gestation week. Assumed 100% adherence to protocol for diagnosis and treatment. Utilities of adverse events were based on assumptions. Long term costs of adverse events (IVH) were based on incorrect calculations and questionable assumptions
van Baaren 2017[82]	Women with symptoms of preterm labour[1], intact membranes and gestational age 24 - 34 weeks (APOSTEL-I) Netherlands	Societal	Hospital	Treat all vs. fFN	Decision Tree	Neonatal death or hospital discharge	proportion of patients treated, perinatal death, composite of adverse neonatal outcomes (perinatal death, CLD, neonatal sepsis, IVH > grade II, periventricular leucomalacia > grade I, and necrotizing enterocolitis), Costs (healthcare, medical & non-medical transport & indirect)	Cost-effectiveness Cost per death avoided Cost per neo-natal adverse event avoided	Assumptions: full compliance with diagnostic and treatment protocol. Treatment was defined as administration of tocolysis and steroids, “combined with the transfer of women to a perinatal center if they were currently in a general hospital”. Preterm delivery was defined as delivery within 7 days after presentation. It distinguished between women who deliver before 34 weeks gestation and those who deliver after 34 weeks gestation. Accounted for different levels of intensity of care, i.e. admission on medium-,high-, or intensive care wards, and in-utero transfers.

Notes: 1. Contractions (more than 3/30 min), vaginal bleeding, or abdominal or back pain.
Key: CLD, Chronic Lung Disease; fFN, qualitative fetal fibronectin; IVH, intraventricular haemorrhage; LOS, length of stay; NHS, National Health Service; PTL, preterm labour; RDS, Respiratory distress syndrome; SPTB, spontaneous preterm birth; QALY, quality-adjusted life year

Figure 8 Decision tree of fFN testing strategy



Key: fFN, fetal fibronectin

Chuck and Nguyen estimated that the introduction of fFN led to an extra 27 ambulance transfers, one fewer hospital admission, and 143 more hospital days for women who were not in labour, relative to what would have happened had testing not been done, during the 2008-2013 period of observation.[79] There were 69 more ambulance transfers and an additional 1,379 days in hospital among women in premature labour. The costs of these health care resources and the additional testing led to an overall increase in costs of US$4 million.

One limitation of Chuck and Nguyen's study is that it was a retrospective study that relied on administrative coding data to identify cases of preterm labour and preterm birth, which is likely to render estimates of 'real world' test accuracy performance unreliable. Another limitation, also acknowledged by the authors, is in their omission of the costs and benefits associated with fFN testing from additional false negatives and true positives mediated through the increases in the proportion of patients born in tertiary care units. Furthermore, the major limitation of this study from our perspective is the lack of assessment of health outcomes. The strengths of the study are found in its documenting of patient management consequent upon test results, particularly in relation to transfers from lower level to tertiary units, and hospital admissions.

Table 14 Parameter values in Chuck and Nguyen 2015[79]

	Inpatient database			Outpatient database			
	Estimate	95% CI Lower	Upper	Estimate	95% CI Lower	Upper	Source
Not in preterm labour							
Transfers							
Positive test results	0.32	0.26	0.39	0.14	0.11	0.18	Chuck and Nguyen 2015
Probability of transfer if not tested	0.06	0.05	0.07	0.03	0.02	0.04	
OR positive vs. not tested (OR+)	2.22	1.38	3.57	10.81	3.96	19.51	
OR negative vs not tested (OR-)	0.78	0.51	1.19	1.53	0.81	2.88	
OR positive vs. negative	2.85	N/A	N/A	7.06	N/A	N/A	Calculations by AG = OR+/OR-
Hospital admissions							
Positive test results	UA	UA	UA	0.11	0.09	0.12	
Probability of admission not tested	UA	UA	UA	0.11	0.10	0.11	
OR positive vs. not tested (OR+)	UA	UA	UA	5.38	3.65	7.95	
OR negative vs not tested (OR-)	UA	UA	UA	0.47	0.37	0.60	
OR positive vs. negative	UA	UA	UA	11.44	N/A	N/A	
In preterm labour							
Positive test results	0.41	0.35	0.46	0.31	0.28	0.34	Chuck and Nguyen 2015
Probability of transfer -not tested	0.06	0.05	0.07	0.22	0.21	0.23	
OR positive vs. not tested (OR+)	7.45	3.89	14.27	3.68	2.55	5.31	
OR negative vs not tested (OR-)	1.91	1.11	3.29	1.26	0.96	1.66	
OR positive vs. negative	3.90	N/A	N/A	2.92	N/A	N/A	Calculations by AG = OR+/OR-
Hospital admissions	UA	UA	UA	Not reported	Not reported	Not reported	Model assumes no cost savings are realised

Key: AG, assessment group; CI, confidence interval; OR: Odds Ratio; N/A, not applicable. UA, unavailable

A UK study also modelled the cost difference between fFN plus clinical examination with clinical examination alone based on signs and symptoms.[77] Costs were measured for the time of hospital observation up to delivery, as the evaluated test strategies were assumed by the authors not to differ in their neonatal costs and consequences. The model was populated with values for hospital admission rates, incidence of tocolysis use, and incidence of steroid use from a UK RCT data reported by Dutta and Norman (2011)[75] (see Table 15). In terms of costs, the analysis used an activity-weighted average LOS of NHS Reference costs Health Resource Group (HRG) NZ07 and NZ08 (for short and long stay), while the rate of hospital transfers, and the proportion of tocolysis administered intravenously were assumed the same across arms, as was the number of ultrasounds per admission (n=1). Due to lack of data, the price of a fFN pathology-based test was used in this study instead of the intended rapid fFN test, and prices for tocolysis and steroids were obtained from BNF sources and doses from the guidelines of the Royal College for Obstetricians and Gynaecologists. The

costs of hospital transfer and the cost of ultrasound (HRG 501OU) were obtained from NHS Reference costs. The study found the rapid fFN strategy saved hospital costs that were partly offset by an increase in diagnostic test costs, resulting in an overall saving of £23.88 per patient in healthcare costs to the NHS.

Table 15 Model parameter values in Deshpande et al. 2013

Parameter	Value	Standard error	Source
Admission rate with fFN +	1.00	.	Reproduction from Dutta and Norman 2011[75] by Deshpande et al. 2013[77].
Incidence of tocolysis with fFN +	0.286	=(0.286*0.714/7)^0.5	
Incidence of steroids with fFN +	0.714	=(0.714*0.286/7)^0.5	
Transfer from hospital with fFN +	0.167	=(0.167*0.833/6)^0.5	
Admission rate with fFN -	0.324	=(0.324*0.676/37)^0.5	
Incidence of tocolysis with fFN -	0.027	=(0.027*0.973/37)^0.5	
Incidence of steroids with fFN -	0.297	=(0.297*0.703/37)^0.5	
Transfer from hospital with fFN -	0.056	=(0.056*0.944/36)^0.5	

Key: fFN fetal fibronectin

5.5.1.2 Cost-effectiveness studies

Boyd et al. 2011 designed a decision tree model with the aim of informing the design of a non-inferiority RCT[83] of fFN that included measuring cost-effectiveness neonatal outcomes from the NHS perspective.[80] The model measured the benefits of accurately diagnosing pre-term birth with fFN and treating with steroids, and the costs of false negative test results in terms of neonatal mortality and morbidity. These were measured relative to the status quo at the time, which was clinical examination and an 'admit all approach'. Unlike other models in this field, the authors of this model assigned a less than 100% admission probability (93%) given positive test results, based on UK audit data [84, 85] They also adopted a 90% probability of admission in the clinical examination only arm, as a best guess assumption (Table 16). The model included the costs of hospital transfers, in addition to those of hospitalisations. On the other hand, it omitted outcomes in terms of inadequate steroid use (i.e. outside the 48 hour to 7 day window before delivery) due to false positive test results (Boyd et al. 2011 Table 16).[80] Further, the model does not account for variation in costs and benefits by gestational age, thus ignoring the dramatically different implications of missing e.g. a premature births under 28 weeks vs. other groups. Thirdly, it did not measure negative effects of steroids use in false positive cases and assumed that only pre-term infants who received intensive or specialised care, 24% (in the ORACLE II RCT, Kenyon et al. 2001)[86], are exposed to mortality risks. The study conclusions were that fFN saved costs but had a "small but potentially detrimental" increase in neonatal morbidity, and a "negligible increase in mortality"[80]

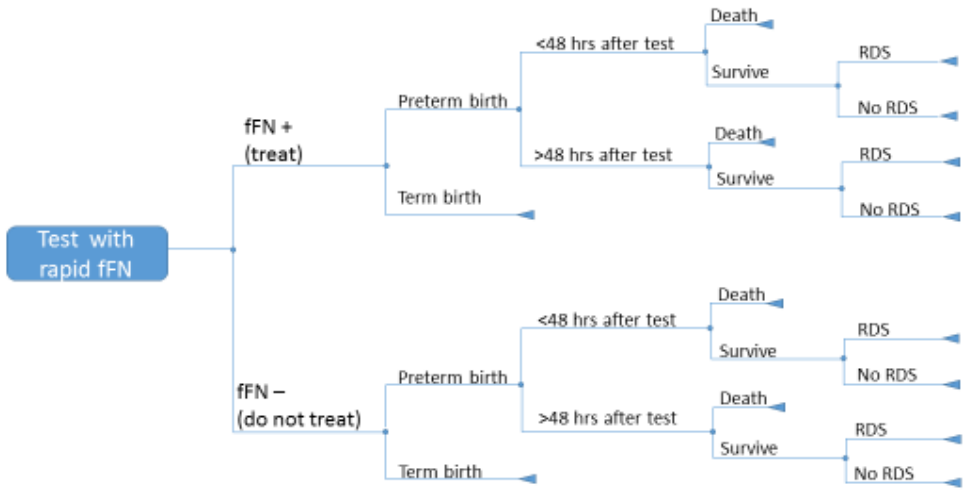
Table 16 Key parameters from Boyd et al. 2011[80]

Model Parameter	Value	Source and comments
Probability of preterm birth	0.20	Probability of PTB among TPL population in ORACLE II (Kenyon et al. 2011); however, the figure is not found in the source.
Probability of preterm morbidity	0.244	Admission to neonatal intensive care (IC) or specialised care (SC) (equivalent to BAPM level 1-3) in ORACLE II (Kenyon et al. 2011).
Steroid risk reduction	0.54	Relative risk preterm morbidity reduction (i.e. admission to IC or SC care) with steroids (Roberts and Dalziel 2006)
Probability of death	0.0257	Probability of death in preterm births (ISD 2008)
Probability of hospital admission with fFN +	0.93	Audit data (HC 2008, Hogg, Penney and Carmichael 2007)
Probability of hospital admission with fFN -	0.33	Audit data (HC 2008, Hogg, Penney and Carmichael 2007)
Probability of admission with clinical examination strategy	0.90	Assumption
Risk of hospital transfer	0.35	Risk of transfer of admitted women to another hospital (Macintyre-Beon et al. 2007)
Cost of hospital admission	£1068	Maternity inpatient cost per stay (average 2.2 days) including drug or treatment
Cost of hospital transfer	£1000	Cost to the NHS of transfer between different hospital CLUs. Value is based on assumption.

Key: BAPM, British Association of Perinatal Medicine; CLUs, consultant led units; fFN, fetal fibronectin; PTB, preterm birth; TPL, threatened preterm labour;

A US evaluation of the rapid fFN and the traditional fFN (treat all with tocolysis and steroids for 24 hours while awaiting the test results) found that the former was more costly and led to more RDS cases and more deaths than the latter.[81] The study compared these strategies against the strategy of treating all pregnant women with steroids as outpatients, which had an incremental cost per RDS avoided of US$433,000, and a cost per neonatal life saved of US$1,300,000, using 1999 prices. A novel feature of this evaluation was its account of adequate corticosteroid administration in the causal chain from testing to neonatal outcomes, through explicit modelling of preterm birth within 48 hours of testing (Figure 9). The model was specified by 1) estimating the probability of premature delivery (before 37 weeks), 2) estimating the probability of delivery within 48 hours of testing among those who are destined to deliver prematurely, 3) estimating the effectiveness of tocolysis in delaying delivery beyond 48 hours and applying these estimates to the baseline probability of delivery within 48 hours. Tocolysis was assumed not to affect the probability of preterm delivery, and the model accounted for the reduced effects of tocolysis due to 24-hour treatment as opposed to a 48-hour treatment course. The sensitivity and specificity of rapid fFN in predicting pre-term birth were used to populate the model. Relevant parameters from this analysis are presented in Table 17.

Figure 9 Model of rapid fFN testing from Mozurkewich et al. 2000[81]



The major contribution of this study was the evaluation of diagnostic effects on neonatal outcomes, and the role in these of tocolytic and steroidal treatment. The study measured costs of test administration, hospitalisation and treatment, maternal cost of delivery, and neonatal hospitalisation costs until death or discharge. Although the study was designed to evaluate cost-effectiveness in a tertiary care unit, and consequently did not measure the costs of transfer in utero or acute neonatal transfers, the costs of antenatal transfers of women presenting with TPTL to a lower care unit would be straightforward to incorporate in this model. A limitation of this analysis was assuming that all patients having positive test results would be admitted and treated with tocolytics and steroids and that all of those testing negative would be sent home. In practice some women testing negative may be admitted due to considerations other than the detection of fFN and, less frequently, some women testing positive may be discharged.[80]

Table 17 Effectiveness parameters in Mozurkewich et al. 2000[81]

Parameter	Value	Range	Source
Probability of preterm birth	0.50	0.20-0.71	Gyetvai et al. 1999, Moutquin et al. 1992 1992[87, 88]
Baseline proportion of premature deliveries who are delivered within 48 hours	0.5	0.2-0.8	Moutquin et al. 1992 1992[88]
Effectiveness of tocolytics for delay of birth <48 h	0.44	0.18-0.62	Gyetvai et al. 1999[87]
Fractional decrease in effectiveness of tocolytics with short-term treatment	0.5	0.2-0.8	Authors' assumption (arbitrary due to lack of data)

Probability of RDS	0.23	0.07-0.67	Crowley et al. 1995[89]
Effectiveness of optimal corticosteroids in preventing RDS	0.65	0.54-0.74	
Fractional decrease in effectiveness of corticosteroids with suboptimal treatment	0.36	0-0.80	
Baseline probability of neonatal death	0.11	0-0.26	
Effectiveness of corticosteroids in preventing death	0.34	0.24-0.52	
Fractional decrease in effectiveness of corticosteroids in preventing death with suboptimal treatment	0.65	0.5-0.9	Authors' assumption (based on the evidence of Crowley et al. 1995 on reduced effectiveness in terms of RDS)

Key: RDS, respiratory distress syndrome

A model for the Netherlands used information on treatment effects of tocolytic and steroid administration from the APSOTEL-II study[90] and diagnostic accuracy from APOSTEL-I[91], reviewed in Chapter 2, to evaluate qualitative fFN, against a treat all strategy.[82] The economic evaluation was performed from a societal perspective in different settings, including general hospitals and tertiary care hospitals. Accordingly, costs of transfer borne by the health system and patients varied depending on the level of the hospital of presentation. Indirect costs of productivity losses were also measured. The model measured perinatal death and adverse neonatal outcomes as a composite measure (including perinatal death, chronic lung disease, neonatal sepsis, IVH > grade II, periventricular leucomalacia > grade I, and necrotizing enterocolitis). The model separately measured the outcomes of infants born within 7 days of testing, after 7 days and before 34 weeks, and after 34 weeks of gestation (see Figure 10), to account for the varying effectiveness of treatment (corticosteroids) with time to delivery. Costs and benefits were measured up to neonatal hospital discharge or death, and results were presented in terms of incremental cost per neonatal adverse event avoided and cost per neonatal death avoided. The authors justified their choice of time horizon on the basis that lack of data made projections highly uncertain. Key model parameters are presented in Table 18.

Figure 10 Model of fFN diagnostic testing in van Baaren et al. 2017[82]

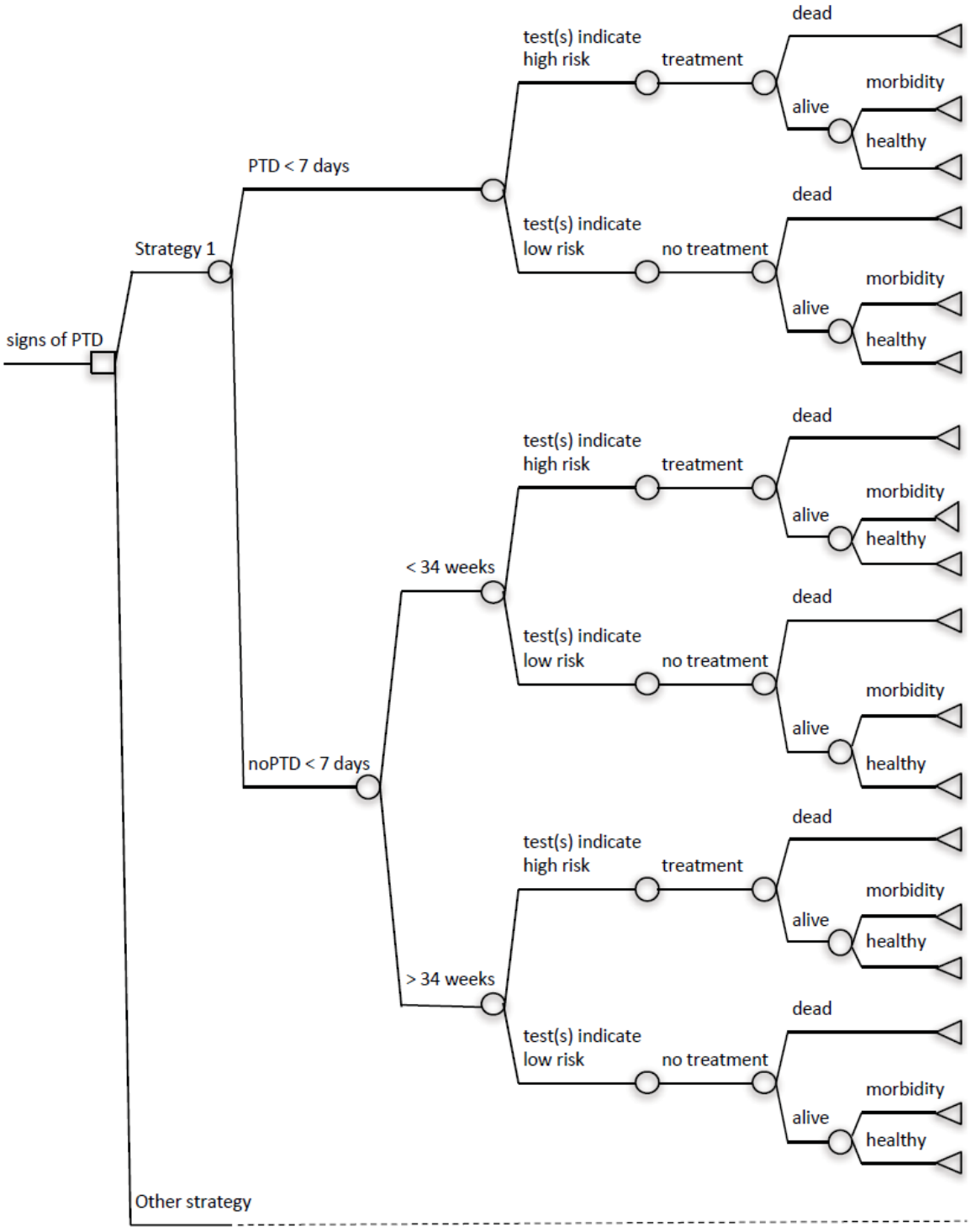


Table 18 Effectiveness parameters in Van Baaren et al. 2017[82]

Parameter	Value	Range	Source
Probability of preterm delivery within 7 days of presentation	0.14	0.12-0.17	Apostel-I, Van Baaren et al. 2014[92]
Preterm delivery after 7 days post-testing and before 34 weeks	0.10	0.08-0.13	Apostel-I, Van Baaren et al. 2014[92]
fFN positive in PTB within 7 days	0.90	0.82-0.94	Apostel-I, Van Baaren et al. 2014[92]
fFN positive in PTB >7 days	0.61	0.49-0.72	Apostel-I, Van Baaren et al. 2014[92]
fFN positive in birth≥34 weeks	0.37	0.33-0.41	Apostel-I, Van Baaren et al. 2014[92]
Perinatal death with antenatal corticosteroids in PTB within 7 days		0.02-0.10	Apostel-II, Roos et al. 2013[90]
Perinatal death with antenatal corticosteroids in PTB after 7 days	0.04	0.02-0.08	Apostel-II, Roos et al. 2013[90]
Perinatal death in births ≥34 weeks	0.01	0.00-0.02	Apostel-II, Roos et al. 2013[90]
Severe adverse neonatal outcomes[1] with antenatal corticosteroids in PTB within 7 days	0.29	0.22-0.38	Apostel-II, Roos et al. 2013[90]
Severe adverse neonatal outcomes[1] with antenatal corticosteroids in PTB after 7 days	0.19	0.14-0.26	Apostel-II, Roos et al. 2013[90]
Severe adverse neonatal outcomes[1] in births ≥34 weeks	0.01	0.00-0.02	Apostel-II, Roos et al. 2013[90]
RR of perinatal death with corticosteroids	0.77	0.67-0.89	Roberts and Dalziel 2006[35]
RR of severe adverse neonatal outcome[1] with corticosteroids within 7 days [2]	0.59	0.41-0.88	Roberts and Dalziel 2006[35]

Note: 1 Composite of Intraventricular haemorrhage >grade II, Chronic Lung Disease, periventricular leukomalacia >grade I, necrotising enterocolitis, neonatal sepsis and perinatal mortality 2, The reporting of these parameters in Van Baaren et al. 2017 appears inconsistent, as no RR is presented for adverse neonatal events with corticosteroids in PTB (<34 weeks) after 7 days.

Key: fFN, fetal fibronectin; PTB, preterm birth; RR, risk reduction

5.5.1.2.1 The NICE 2015 Guidelines model

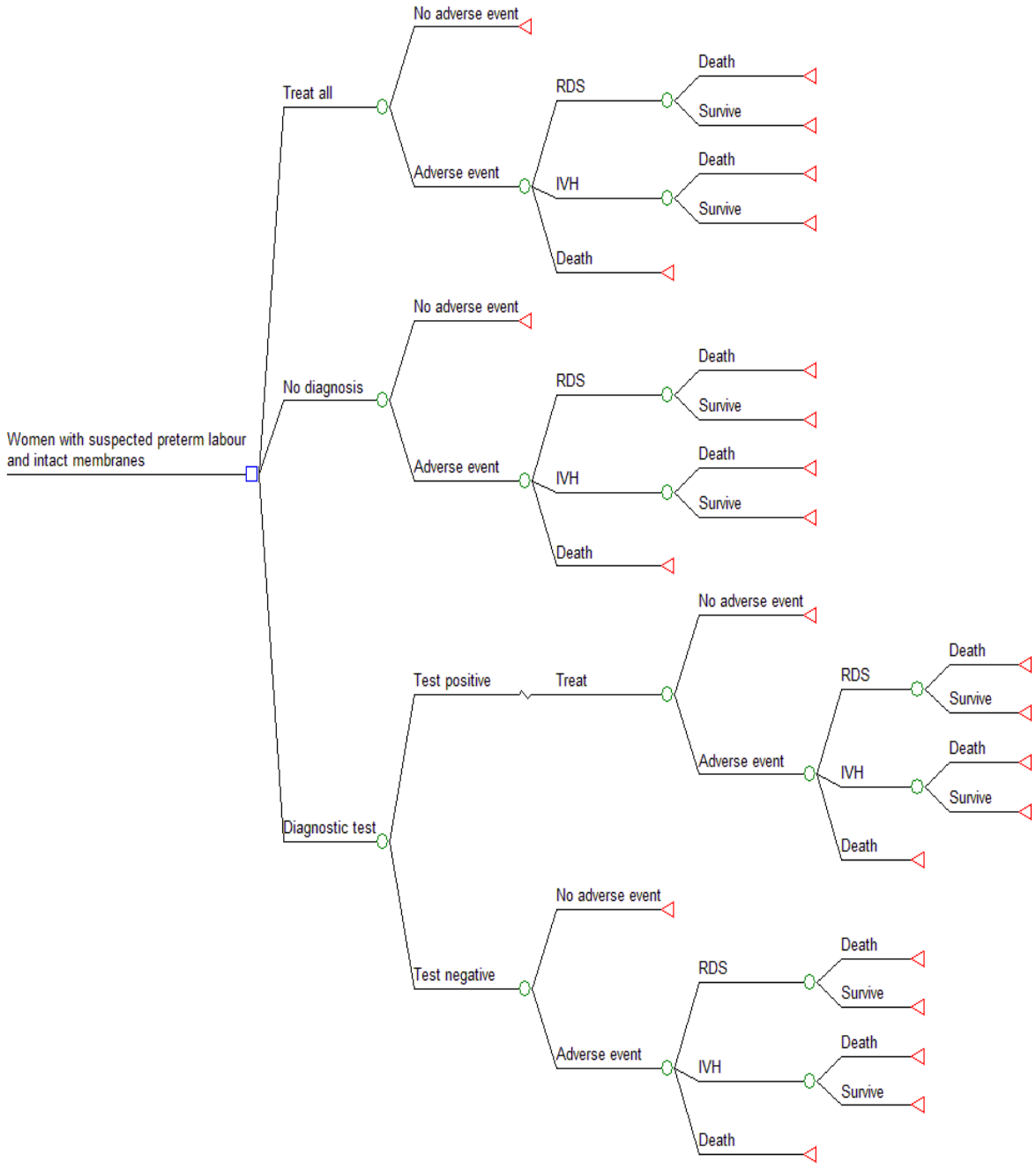
The Royal College of Obstetrics and Gynaecology (RCOG) developed a decision analytic model to inform the NICE 2015 guidelines on the diagnosis and treatment of preterm labour and birth.[26] The authors of this model concluded that the quality of the diagnostic accuracy data was low for the different tests considered relevant at the time for women presenting with symptoms of preterm labour (cervical length measurement by ultrasound, Actim Partus, qualitative fFN) was low. Consequently they presented a 'what if' analysis comparing the testing vs no testing-treat all strategies, consisting of identifying the levels of specificity and sensitivity at which a hypothetical test became cost-effective according to the NICE cost-effectiveness threshold of £20,000 per QALY gained.

Unlike previous analyses, the NICE evaluation accounted for the effect of gestational age on the trade-off between sensitivity and specificity, i.e. costs of treating more patients unnecessarily versus missing patients at high risk of neonatal adverse events, including death. This analysis set the cost of the test equal to that of cervical length measurement, and found that testing was not cost effective for gestational ages below 30 weeks. This

served as the basis of the NICE recommendations about use of testing to rule out preterm labour.

The NICE guidelines model structure is illustrated in Figure 11. According to this model, the causal pathway from diagnostic results to neonatal outcomes is mediated by tocolysis treatment, which can delay premature delivery by 48 hours or more. This would generate a window of opportunity for appropriate steroid administration (i.e. at least 24 hours and up to 7 days before delivery) and transfer to a tertiary hospital, thus reducing the risk of RDS, IVH, and death. The risk reduction parameter values used in this model are based on treatment effect estimates for calcium channel blockers vs. placebo, from three separate network meta-analyses of RCTs, one per model outcome (NICE 2015, Chapter 10, p. 184-238).[26] It is worth noting that two of these treatment effect parameters, the odds ratio (OR) for death 0.62 (95% CI: 0.21 to 1.80) and for RDS 0.81 (95% CI: 0.50 to 1.34), are imprecisely estimated; the estimated OR for IVH was 0.40 (95% CI: 0.21 to 0.74); none of these estimates used data from a direct head-to-head RCT. Furthermore, these treatment effects were assumed to be constant across the gestational ages in the model (24 to 34 weeks), so that the absolute risk reduction (ARR= relative risk x baseline risk) for the three types of event with tocolysis varies by gestational age only because of the baseline (i.e. without tocolysis) risk declines with gestational age (See Figure 12 and Table 12). IVH and RDS each contribute to the risk of neonatal mortality; the probability of death conditional on the former event is 0.300 and is 0.054 conditional on the latter; infants who did not die following these adverse events would contribute additional costs and QALY losses over their expected lifetime.

Figure 11 The NICE 2015 Guidelines model[26]



The analysis projected life-long QALY values based on the neonatal adverse events. Infants who survived the neonatal phase and were discharged home were assumed to have an average expected QALY that varied with gestational age from 19.92 at 24 weeks to 22.61 at 34 weeks. These were calculated as the gestational age-specific proportion of infants surviving the first year of life times the expected QALYs of these infants, which in turn was equal to the life expectancy in England and Wales of 80 years, valued at the population norm health state utility of 0.82 and discounted at an annual rate of 3.5%.[93] Deaths in the first year of life were assumed to generate 0 QALYs. In the event of RDS, an average QALY

loss would apply, arbitrarily set at 3.85 (based upon RDS providing a slightly lower QALY loss than IVH). The occurrence of IVH incurred an average QALY loss of 4.5, based on the assumption that the IVH would incur the same quality of life loss as intracranial haemorrhage (ICH), which in turn was assumed to incur one third of the QALY loss of cerebral palsy value reported by Cahill (Cahill 2011).[94]

The analysis accounted for the costs of tests, including test acquisition and staff time, drug treatment (nifedipine at loading dose of 40 mg, and subsequent dose of 240 mg; at British National Formulary (BNF) prices of £0.008 per mg, from a 90 capsule pack of 10 mg) and administration (5 minutes of doctor and 5 minutes of nurse time), and downstream neonatal hospital costs of adverse events (RDS and IVH). The downstream RDS costs were set at the NHS Reference costs of NICU care (BAPM level 1) with ECLS/ECMO (extracorporeal life support or membrane oxygenation), while the costs of IVH were assumed to be equal to the lifetime healthcare costs of ICH, which in turn were assumed to be equal to the healthcare costs of severity Grade III or Grade IV cerebral palsy. Regardless of how valid these clinical assumptions are, the cost of IVH appears to be under-estimated, as it was calculated with a higher discount rate than the 3.5% recommended by NICE, and underweighted:

> *"It was additionally assumed that Grade III and Grade IV ICH would be similar in cost to cerebral palsy. A European paper[95] estimated in year 2000 prices that the lifetime healthcare costs for cerebral palsy using an annual discount rate of 5% was €66,155 for men and €65,288 [for women]. The mid-point of this estimate was used and converted into GBP using an exchange rate of £0.83 = €1 … It was then converted into 2011/12 prices using the HCHS (The Hospital & Community Health Services) Index. One study [96] suggested that 30% of ICH is of severity Grade III and Grade IV and therefore the cost of ICH was estimated as 0.3 x £79,000"* (NICE 2015, footnote c to Table 129, p. 406)[26]

The parameter estimates for the different elements of cost appeared to be estimated in prices of different years. Treatment costs were expressed in 2015 prices, costs of drug administration were in a price year prior to 2014, adverse events were in 2011/2012 (IVH) and 2012/2013 (RDS) prices. The long term adverse event cost (IVH) was derived from a European study that reported results in 2010 €, converted to UK £ using the 2014 exchange rate, and reflated to 2011/12 prices.

The key assumptions of the NICE Guidelines model are summarised in Table 19. As in other models in this literature, maternal outcomes are not measured. The model assumes full adherence to the diagnostic protocol, thus abstracting from individual disparities in clinician behaviour. Critically, the model assumes that all patients with positive test results are treated

with tocolytics, which may not occur in routine practice (Andrew Shennan personal communication August 2017). The model does not explicitly account for the use and effect of corticosteroids, only implicitly within the treatment effect estimates obtained from the network meta-analysis of tocolysis studies discussed above. This feature makes the model less suitable for obtaining generalisable results for situations where the corticosteroids are used without tocolytic therapies. A major limitation as discussed above is the high degree of uncertainty associated with the calculation of costs and QALYs, which were extrapolated to lifetime from neonatal morbidity outcomes. Thus this model's advantage in terms of producing results in terms of QALYs for informing NICE decisions may have come at the cost of heroic assumptions about the ability to predict lifetime costs and benefits from neonatal outcomes. In fact the extrapolation was inadequately calculated (by multiplying a life expectancy times a constant population norm) since it did not account for the survival curve profile in population life tables and the varying utility with age (Ara and Brazier 2010), and the utility norms were derived from a study that predates the time EQ-5D scores were developed. Also it is unclear whether by choosing to model treatment based on tocolysis as opposed to steroids, the model failed to account for outcomes in terms of other neonatal adverse events, such as necrotising enterocolitis, sepsis and retinopathy. On the other hand, the model's ability to account for outcomes by gestational age at presentation make this model the most relevant among those available for guiding clinical decisions on individual patients, since other models did not produce results by gestational age. A summary of the main features of the model is provided using the CHEERS (Husereau et al. 2009) checklist in Table 20.[69]

5.6 Discussion and further research

There is room for improvement in the parameter values used to populate the model, particularly in terms of long term disutility values of adverse events (intraventricular haemorrhage and respiratory distress syndrome) and the health utility population norm used for preterm survivors, which is outdated. Furthermore the extrapolation of utility values does not account for survival curves in life tables. In terms of costs, the quality of data collected for the NHS Reference costs of critical care BAPM levels 1 -4 is low as it is unlikely equate with the actual costs of care but typically reported by hospitals without adequate apportioning by level of care (Eleri Adams, personal communication, 09 June 2017); on the basis of consulted expert advice, we propose NHS tariffs for the four levels of care may provide better estimates of true economic cost. Although these may still be biased due to factors other than costs, such incentives to reduce costs, the may be less biased than reference costs. Further, the costs of IVH were based on inappropriate calculations and data

for another disease, and systematic searching for the literature for better estimates of this cost parameter seems worthwhile.

Other areas of uncertainty that deserve to be explored include the following:

- Mortality benefits of reducing the rate of false negative cases, who, depending on the nature of local hospital (i.e. a level 2/3 vs level 1 hospital), may be at increased risk of mortality if birth occurs before 32 weeks gestation (Mujica-Mota et al. unpublished; see next chapter)
- Accounting for differences in treatment costs of positive cases according to level of hospital of presentation, due to the costs of in-utero transfers for very preterm pregnancies as well as their repatriation to the local hospital after neonatal stabilisation.
- Accounting for compliance with the treatment protocol subsequent to diagnostic test findings

In the next sections we undertake these revisions to the NICE Guideline model and populate them with the diagnostic test accuracy evidence from the systematic review from Section 2.

Table 19 Key assumptions in the NICE 2015 Guidelines model

Assumption	Description	Critique
i) The choice of diagnostic strategy has no clinically and economic significant effect on the mother	The clinical outcomes, costs and QALYs associated with the mother are not measured	Implicit is the view that the outcomes of the mother are either irrelevant for the policy maker's decision on how to diagnose preterm labour.
ii) Full adherence to the diagnostic protocol	All individuals testing positive are admitted to hospital and given treatment	Audit data from England found that 7% of patients testing fFN positive were not admitted and 32% testing fFN negative were admitted (Healthcare Commission 2008; Hogg, Penney and Carmichael 2007);
iii) the effects of diagnostic testing on neonatal outcomes are mediated through treatment with tocolytics	All individuals are treated with tocolytics	Tocolytics is now being used infrequently (Andrew Shennan, personal communication August 2017)
iv) steroid use is not explicitly modelled but implicit in the tocolytic treatment effect values estimated from the literature;	Tocolytics may be given to postpone delivery for at least 48 hours to a) allow in-utero transfers and/or b) treat with steroids	Some protocols on the use of quantitative fFN, e.g. London's St. Guy's and Thomas, provide different guidelines for the decision to admit and the decision to treat with tocolytics and to treat with steroids
v) the relative effect of tocolytics are constant across gestational ages	Tocolysis reduces the risk of adverse neonatal outcomes, including death, at a constant proportion across gestational ages	This is an untested assumption driven by the available data
vi) neonatal morbidity outcomes are measured in terms of RDS and IVH;	RDS and IVH are two of the key outcomes reported in the evaluation literature on tocolytics treatment	The network meta-analysis evidence used to populate the treatment effects on these outcomes is based entirely on indirect comparisons and the treatment effect estimates are consistent with no effect (i.e. Odds ratio credible intervals cross 1).
vii) neonatal mortality may occur through the risk of death associated with RDS or IVH or	The effect of tocolysis on neonatal mortality is divided between an indirect effect,	The network meta-analysis evidence used to populate the treatment effects on this outcome was based on no head to head data on tocolysis

background risks that decline with gestational age	operated through its effect on RDS and IVH, and a direct effect through other causes.	vs no treatment and the treatment effect estimate is consistent with no effect (i.e. Odds ratio credible interval crosses 1).
viii) The expected lifetime quality of life of infants who survive the first year after birth without RDS or IVH is the same for full term and preterm infants	Conditional on surviving the first year of life and neonatal morbidity outcomes, lifetime QALYs are independent of gestational age at birth	This assumption is questionable in the light of evidence of long-term health and behavioural problems associated with pre-term birth
ix) The expected lifetime costs for preterm and full term infants who survive the first year without IVH are the same	Conditional on surviving the first year of life and IVH occurrence, lifetime costs are independent of gestational age at birth	This assumption is questionable in the light of evidence of long-term health problems associated with pre-term birth

Key: fFN, fetal fibronectin; IVH, intraventricular haemorrhage; QALY, quality-adjusted life year; RDS, respiratory distress syndrome

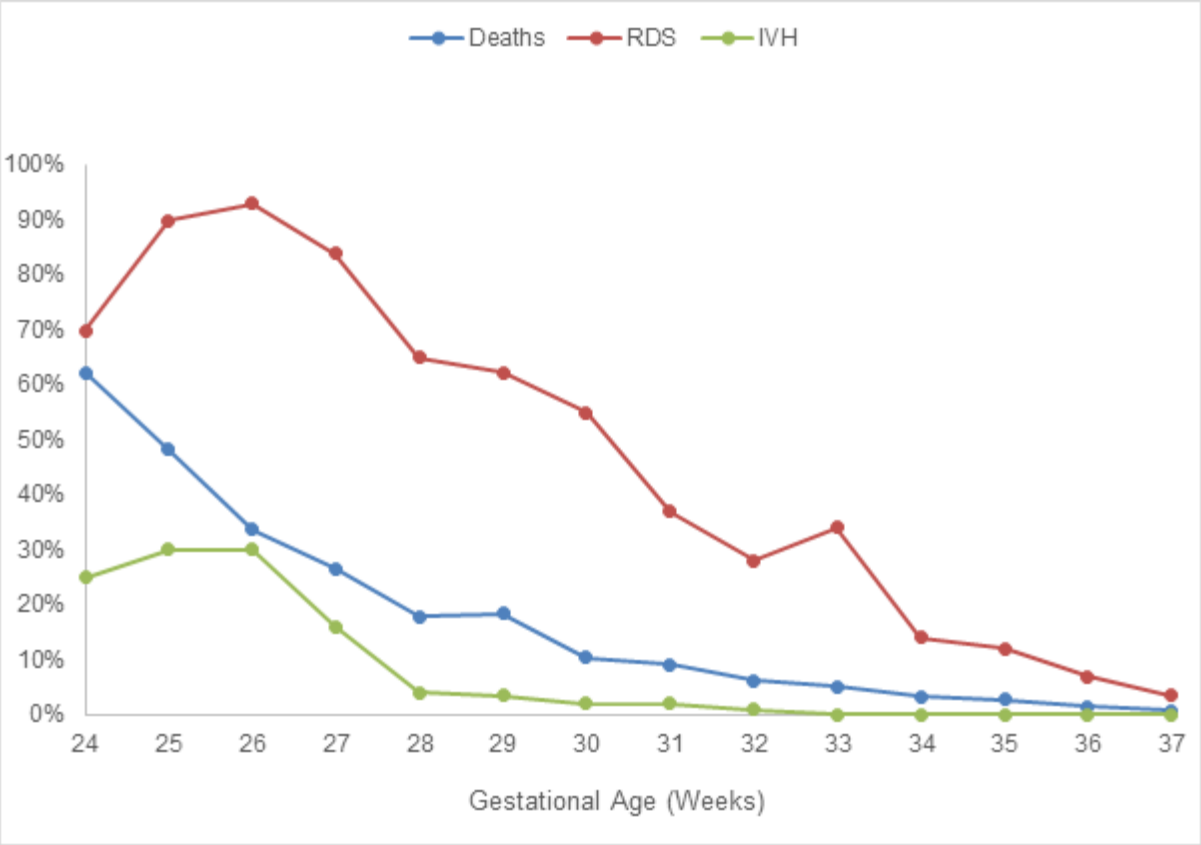
Table 20 CHEERS checklist for the NICE Guidelines Model [69]

Item	Item no.	Recommendation	Reported on page no.
Methods			
Target population and subgroups	1	Describe characteristics of the base case population and subgroups analysed including why they were chosen	Diagnosis of preterm labour in women with intact membranes presenting with symptoms suggestive of pre-term labour Section 9.6, p. 176-177, and Section 1.3, p. 350 of NICE 2015 Guideline (NICE 2015)
Setting and location	2	State relevant aspects of the system(s) in which the decision(s) need(s) to be made	The model does not account for cost of in-utero transfers, thereby implicitly assuming that women presents to a level 3 hospital.
Study perspective	3	Describe the perspective of the study and relate this to the costs being evaluated	The NHS perspective was adopted. Healthcare costs are based on NHS Reference cost sources, and costs of medications are form British National Formulary (BNF) prices
Comparators	4	Describe the interventions or strategies being compared and why they were chosen	It compared testing vs. no testing-treat all vs. no test and no treat at different gestation ages, to derive the thresholds of sensitivity and specificity that would make testing cost-effective. This 'What if' assessment was conducted in the light of low quality of diagnostic test accuracy data
Time horizon	5	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate	Lifetime of child, based on imputed long term costs and QALYs on the basis of neonatal adverse events. The life time horizon is appropriate since neonatal outcomes on which the choice of strategy impact (respiratory and cognitive) have long term quality of life and resource need implications.
Discount rate	6	Report the choice of discount rate(s) used for costs and outcome(s) and say why appropriate	3.5% for both costs and QALYs, as recommended by the NICE reference case.
Choice of health outcomes	7	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis undertaken	QALY. This combines morbidity and mortality outcomes in a single index measure for comparison across disease areas, as required for informing NICE decisions.
Measurement of effectiveness	8	Describe fully the methods used for the identification of included studies and synthesis of clinical data	The NICE guideline model systematically searched for studies of test accuracy of biochemical test, cervical length measurement by ultrasound and clinical examination. However it found that the identified studies were of low quality. The NICE Guideline analysis updated a systematic review comparing tocolytic treatment classes using Network meta-analysis. This method allowed to compare studies who were not investigated directly in any RCT, thus expanding the evidence base for informing the analysis.
Measurement and valuation of preference based outcomes	9	If applicable describe the population and methods used to elicit preferences for outcomes	The expected QALYs at birth for an infant without adverse neonatal events (RDS or IVH), was calculated as the result of the life expectancy at birth of 80 years in England and Wales and this was multiplied by the population utility norms of 0.82 (the details of the citation given for this value, 'Kind 1983', could not be found). The disutility associated with RDS was based on an arbitrary assumption. The disutility associated

			with IVH was based on 1/3 of the utility loss from cerebral palsy.
Estimating resources and costs	10	Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Included were the costs of tests, drug treatment (at BNF prices) and administration (doctor and nurse time), and downstream neonatal hospital costs of adverse events (RDS and IVH). The downstream RDS costs were set at the NHS reference costs of NICU care (BAPM level 1), while the costs of IVH were assumed to be equal to the lifetime healthcare costs of ICH, which in turn were assumed to be equal to the healthcare costs of severity Grade III or Grade IV cerebral palsy. The calculations used in the model appear to underestimate the long term costs of IVH.
Currency, price date, and conversion	11	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	The parameter estimates for the different elements of cost appeared to be estimated in prices of different years: treatment costs were expressed in 2015 prices, costs of drug administration were in a price year prior to 2014, and adverse events were in 2011/2012 (IVH) and 2012/2013 (RDS) prices. The long term adverse event cost (IVH) was derived from a European study that reported results in Euros of 2010, converted to UK£ using an exchange rate of 2014 and reflated them to 2011/12.
Choice of model	12	Describe and give reasons for the specific type of decision-analytic model used. Providing a figure to show model structure is strongly recommended.	In line with prior modelling work, the NICE Guideline used a decision tree model, with long term QALY and costs pay-offs. This is reasonable given the limited amount of neonatal outcome data on which to base modelling of medium to long term outcomes.
Assumptions		Describe all structural or other assumptions underpinning the decision-analytic model.	The model implies the assumption that the mother is unaffected by the diagnostic strategies (no maternal outcomes were measured). It also assumed: i) Full adherence to the diagnostic protocol; ii) the effects of diagnostic testing on neonatal outcomes are mediated through treatment with tocolytics; iii) steroid use is not explicitly modelled but implicit in the tocolytic treatment effect values estimated from the literature; iv) the relative effect of tocolytics are constant across gestational ages; v) neonatal morbidity outcomes are measured in terms of RDS and IVH; vi) conditional on surviving the first year of life and neonatal morbidity outcomes, lifetime QALYs are independent of gestational age at birth
Analytic methods		Describe all analytic methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (e.g., half-cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Network meta-analysis was used to obtain evidence of treatment effects of tocolytic vs.no treatment on mortality and morbidity (RDS and IVH) events. The model used no long term extrapolation; it simply projected costs based on neonatal morbidity and 12-month infant survival after birth.

Key: NHS, National Health Service; QALY, quality-adjusted life year; RDS, respiratory distress syndrome; IVH, intraventricular haemorrhage; BAPM, British Association of Perinatal Medicine; NICU, neonatal intensive care unit.

Figure 12 Baseline (without tocolytics) risks in the NICE Guideline model



Source: (NICE 2015)[26]

6 Independent economic assessment

6.1 Methods

This chapter presents a de novo evaluation of PartoSure, Actim Partus and fFN at thresholds other than 50 ng/ml, relative to fFN at 50 ng/ml as the comparator.

6.1.1 Model structure

In common with all previous studies (discussed in the review of economic studies section), we used a decision tree to model the economic evaluation of the diagnostic choice problem. As in the model supporting the 2015 NICE Guidelines on diagnosis and treatment of preterm labour the only case that included both of these aspects of patient management, our model includes an initial diagnostic phase followed by treatment and long term outcomes.[26] The model accounts for the costs incurred starting from the time women present to a maternity hospital with symptoms suggestive of preterm labour, through hospital admission or discharge home, to neonatal discharge or death in hospital. The health consequences to the offspring are measured in terms of quality adjusted life years (QALYs) based on neonatal morbidity and mortality outcomes. The main features of our model are that it:

- Accounts for the costs and QALYs of the new-born (as well as QALYs for the mother, in a scenario analysis)
- Differentiates costs and benefits by gestational age
- Distinguishes between hospital levels: tertiary (also known as level 3 – capable of dealing with the most severe cases), level 2, or level 1
- Accounts for the costs and benefits of steroids, and the costs of tocolysis and hospital transfer for neonatal transfers
- Determines long term QALYs and costs by neonatal morbidity (RDS and IVH) and mortality outcomes

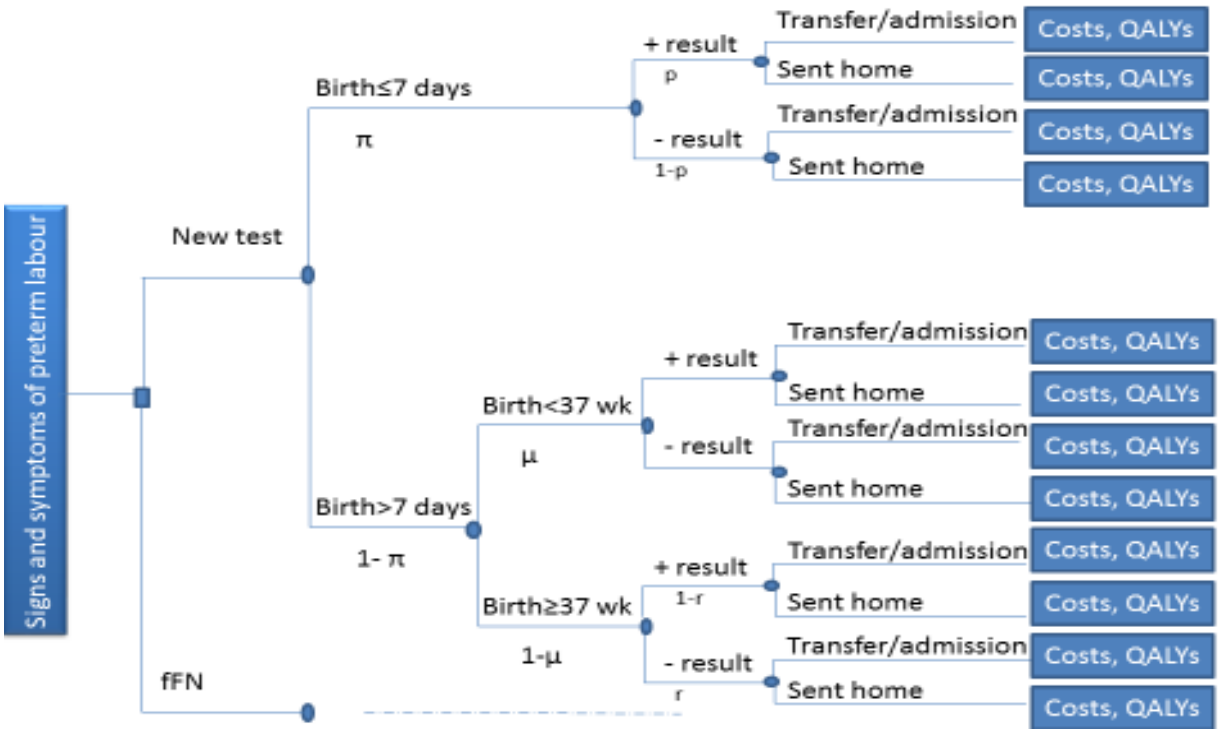
The model builds on that used to inform the NICE 2015 Guidelines on the diagnosis and treatment of preterm labour, as the only prior model allowing for variation in health risks, and thus costs and benefits, of inaccurate diagnosis by gestational age.[26] By adopting this general structure we are able to account for the increasing neonatal health risks posed by an attending obstetrician's failure to identify a woman in preterm labour at earlier gestational ages. Unlike the NICE model, which assumed the diagnosis of preterm labour was intended to guide the decision of whether to administer tocolysis, we model the treatment pathway following a diagnosis of preterm labour around the decision of whether to treat with corticosteroids and/or admit to hospital or discharge home. This methodological variation in

our approach is motivated and informed by the very recent evidence quantifying the positive effects of antenatal corticosteroid (ANS) administration for accelerating the maturation of the foetus' lungs as a function of the time of administration relative to delivery; the limited use of tocolysis reported in the literature; and the emerging consensus on its potential risks to the new-born baby and side-effects to the mother [33, 36, 37, 97] For example, audit data for the period September 2016-May 2017 from the level 2 hospital in Exeter shows that tocolysis was administered to only one out of 9 (11%) cases presenting with symptoms of preterm labour at 24-34 weeks with fFN≥50 ng/ml. In our analysis we assume tocolysis is only used for all in-utero transfers at gestational ages less than 28 weeks (see below in this section).

Given the importance of timing of ANS administration, the aim was to develop a model capable of accounting for the different diagnostic test options' capabilities to distinguish between those likely to deliver imminently following presentation and those who would deliver preterm in a week or later, among presumptive cases of preterm labour. Despite our initial aim to explicitly model treatment administration at intervals of <2 days, 2-7 days and 7 and more days before delivery, few studies of diagnostic test accuracy reported outcomes for the ≤2 days' time-to-delivery interval (see Section 2.2). These intervals have been discussed in the literature as most relevant to ANS effectiveness, with 7 days before delivery considered the earliest time for effective use of steroids in terms of fetal and neonatal mortality and respiratory distress syndrome.[98, 99] The latest evidence suggests that the effectiveness of ANS in terms of mortality risk reduction may be optimal within 2 days to delivery, and that it diminishes with time before delivery. This observation also applies to the risk of respiratory distress syndrome (RDS) and intraventricular haemorrhage (IVH).[36, 97] We thus decided on the model structure illustrated in Figure 13.

This structure shares the features of previous models of diagnosis and treatment and is determined by the a) available diagnostic test accuracy data and b) the latest evidence on the time window relative to delivery when ANS treatment is most effective. The model assumes that the decision to admit and treat or transfer to another hospital is driven by the test result (positive or negative). The model makes a distinction in terms of effectiveness between diagnostic tests according to their ability to correctly predict whether a woman will deliver before term, and whether that will occur before or after 7 days.

Figure 13 PenTAG model structure



Note: dotted line indicates that the subsequent branch structure after 'fFN' is identical to the structure shown for 'New test'. New test is patient management according to one of the interventions or 'index tests'. fFN is the comparator (admit and treat when fetal fibronectin≥50 ng/ml) status quo. We also consider the no-test-treat-all comparator. Greek and Latin letters are parameters populated with data from diagnostic accuracy test studies.

Keys: π = pretest probability of preterm birth (PTB) ≤7 days, p = test sensitivity of PTB ≤7 days, μ= pretest probability of PTB, r = test specificity of PTB.

In this model, the costs and health benefits of following one of at least two mutually exclusive courses of action for managing a woman presenting with signs and symptoms of preterm labour are evaluated. Figure 13 shows that a new test (PartoSure, Actim Partus, quantitative fFN used qualitatively at thresholds other than 50 ng/ml) or the no-test treat-all strategy may be compared against the status quo of qualitative fFN (or quantitative fFN used qualitatively at a 50 ng/ml threshold). The starting point of the model is when the decision between diagnostic strategies is made, immediately after clinical assessment of symptoms that have not ruled out preterm labour. Thus, in the absence of further testing, all women would be admitted or transferred to another hospital. A woman tested may turn out to deliver a preterm baby within 7 days of testing or may deliver in more than 7 days from the time of testing. If the latter occurs, birth may be preterm (before 37 weeks' gestation) or full term. Women may therefore be classified in one of these three subgroups according to the time of delivery. Within each of these, the results of the new test will determine how the patient is managed, and consequently the woman's ability to benefit from ANS treatment. Thus, if a woman tests positive, the obstetrician would be expected to treat her by admitting her to the

hospital and administering steroids (under fFN testing, in some hospitals women may be admitted for observation above one threshold and admitted and administered steroids at another higher threshold; we do not consider this case). The model distinguishes the type of hospital setting by level of specialisation: if a woman attends a tertiary level hospital at less than 28 weeks' gestation and tests positive, she is admitted into hospital, whereas if testing takes place at a lower level hospital, she would be given tocolysis (we assume tocolysis is only considered for women undergoing in-utero transfer at <28 weeks' gestation) and transferred to a level three hospital for their care. Women who test negative are sent home without treatment; due to lack of any test-specific data on this parameter we do not allow for partial compliance with treatment guidelines contrary to what is suggested in Figure 1. The same structure is assumed for the status quo 'fFN' testing option with one and the same threshold for admission and treatment.

According to Figure 13, a symptomatic woman who goes on to deliver within 7 days has a positive test result with probability p (the sensitivity of the test), and a negative test result with probability 1-p (the false negative rate). Among women who deliver after 7 days, the probability of a positive test result is equal to the false positive rate (FPR), and the probability of a negative test result equals the test specificity for delivery within 7 days of testing.

Some women have a positive test result, receive treatment and deliver after 7 days of testing, but before 37 weeks of gestation. (In the base case analysis we assume that ANS produces no benefit when administered more than 7 days before pre-term birth (Watson, Ridout and Shennan 2015, 2016).[98, 99] In scenario analyses, these women are assumed to benefit from ANS, but less so than those who are treated within 7 days of preterm delivery[36, 97] (throughout we assume no multiple courses of steroids are given, based on obstetricians' advice on routine practice and the perceived lack of proven benefit and risks to neonate). The frequency of such cases is calculated as

$$P(PTB > 7\ days\ \&\ +ve\ result) = (1-\pi) \times (1 - Specificity_{7d}) - (1-\rho_{37w}) \times (1-r)$$

where P(PTB>7days and positive result) is the probability of having a positive test result and delivering preterm > 7 days after testing, ππis the incidence of delivery within 7 days, $Specificity_{7d}$ is the test specificity for delivery ≤7 days, and r is the test specificity for delivery <37 weeks (i.e. the proportion of women testing negative among those who deliver after 37 weeks), which has an incidence of ρ_{37w}. Thus the proportion of women who receive treatment more than 7 days before PTB (and therefore derive partial benefit from ANS) is equal to the difference between the FPR for delivery ≤7 days and the FPR for delivery <37 weeks, weighted by their respective incidences. The benefit from ANS for this group of women is also reduced by the fact that the baseline mortality and adverse event (IVH and

RDS) risk of preterm birth is lower, as the infant is delivered at an older gestational age, than mothers who deliver within 7 days. We assume that delivery takes place at the midpoint between gestational age at presentation and 36 weeks.

6.1.2 Population

The population was defined as: women presenting with symptoms of threatened preterm labour (abdominal pain, contractions) with intact membranes between 24 and 36 weeks' gestation, for whom transvaginal ultrasound is not available or acceptable.

6.1.3 Interventions and comparators

We evaluated the following diagnostic test strategies immediately following an initial clinical investigation that has not ruled out preterm labour:

a) Testing with PartoSure

b) Testing with Actim Partus

c) Testing with quantitative fFN at thresholds of 10, 200 and 500 ng/ml

d) Comparator testing with fFN at 50 ng/ml, from quantitative or qualitative versions of the test device.

e) Treat all without testing; i.e. since clinical investigation could not attribute symptoms to other causes women are managed as presumptive case of preterm labour

These were the options for which evidence was available in the literature. Combinations of these options were not considered, as they were not part of the NICE Scope.

In addition, we explored the scenarios of evaluating a) qualitative fFN according to current treatment protocols in Guy's and St Thomas' women's hospital, where different thresholds are used to admit to hospital (at 50 ng/ml) and treat with steroids (200 ng/ml) and b) quantitative fFN (qfFN) as observed at the level 2 maternity hospital in Exeter.[100] These additional analyses are intended to reflect the spectrum of variation in local current practice across the country.

6.1.4 Perspective, time horizon and discounting

The analysis adopted the perspective of the NHS and Personal Social Services. In accordance with the requirements of the NICE methods guide, the time horizon is taken as the entire lifetime, whereby the projected long term healthcare costs and utilities associated with avoiding an adverse neonatal outcome (death, respiratory distress syndrome and intraventricular haemorrhage) were measured.[101] All previous models of diagnosis in preterm labour assumed much shorter time horizons (up to neonatal death or discharge from

hospital), except for the model informing the NICE 2015 Guideline in this area.[26] (see Chapter 4 for a critique of these models). An annual discount rate of 3.5% for costs and benefits was used, as set by NICE. We present results limited to neonatal death or discharge from hospital in scenario analyses.

6.1.5 Model parameters

The model consists of two parts: the diagnostic phase and the treatment phase, each with a characteristic set of parameters and sources of evidence. The diagnostic phase parameters are populated from diagnostic accuracy studies of the interventions of interest, complemented by data on patient management such as admission rates conditional on test results, which are obtained from audit data or modelling studies. The treatment phase is derived from large observational studies of the effects of steroids on neonatal health outcomes and evidence synthesis of RCTs of antenatal steroids treatment. Cost estimates are obtained from detailed costing studies in individual hospitals or routine national sources, and health related quality of life utilities have been obtained from observational and health state preference elicitation studies.

6.1.5.1 Treatment effectiveness and extrapolation

6.1.5.1.1 Diagnostic test accuracy

We limit our analyses to evaluate the diagnostic tests assessed by individual comparative diagnostic accuracy studies, as identified in the effectiveness section of this report. These studies are the following:

- APOSTEL-1 study comparing Actim Partus with qualitative fFN at the 10, 50, 200 and 500 ng/ml thresholds (Bruijn et al. 2016)[45, 46]
- A comparison of Actim Partus with PartoSure (Hadzi-Lega et al. 2017)[47]

Other studies were identified as providing relevant data but of lower quality than these two studies. One was a comparison of Actim Partus with fFN at 50 ng/ml (Cooper et al. 2012).[1] Cooper and colleagues' was the second largest of the three identified studies that compared Actim Partus vs. fFN at 50 ng/ml and reported test accuracy data for delivery within 7 days and at less than 37 weeks (see Chapter 2.2). However, as discussed in Chapter 2, it was unclear what version of the fFN test was used, but presumed to be the ELISA version of the qualitative fFN test which is no longer used in clinical practice. Therefore we consider this study in scenario analyses, thus limiting the base case analysis to include only the APOSTEL-1 study, which evaluated a non-laboratory based fFN test, and the study by Hadzi-Lega et al. In scenario analyses we also evaluated the comparison of Actim Partus vs. fFN at 50 ng/ml (non-ELISA) tests based on a meta-analysis of 7-day results reported by

four comparative studies of these technologies.[45, 46, 53, 58, 59] We did not consider this meta-analysis in the base case due to the heterogeneity between the combined studies, especially in terms of preterm birth rates, which likely drove their differences in test accuracy results. Furthermore, the pooled results for each index test presented in section 2.2.6 were not considered in the economic analysis, because comparisons between tests in terms of those results are likely confounded by the heterogeneity between the studies.

In addition, although it was excluded from the effectiveness section because it did not provide published test accuracy data within 7 days, we evaluated the diagnostic test considered in the only UK study (Abbott et al. 2013)[102] using data provided by the study authors for this review (Andrew Shennan personal communication August 2017):

An assessment of the rapid fFN10Q analyser (Hologic) at the qualitative thresholds of 10, 50, 200, and 500 ng/ml;

Table 21 Analyses conducted and their sources

	Comparator / Intervention	**Study Name / Data**				
		Base case analysis		**Scenario Analysis**		
		APOSTEL-1[45, 46]	**Hadzi-Lega et al. 2017[47]**	**Cooper et al. 2012[1]**	**Abbott et al. 2013[102]**	**Meta-analysis[45, 46, 53, 58, 59]**
PartoSure	Intervention		✓			
Actim Partus	Intervention	✓	✓	✓		✓
Rapid fFN 10Q Cassette Kit thresholds other than 50 ng/ml	Intervention	✓			✓	
fFN, threshold 50 ng/ml	Comparator	✓		✓	✓	✓
No test, treat all	Intervention	✓	✓	✓	✓	✓

Key: fFN, fetal fibronectin

The sensitivities and specificities used for these analyses are presented in Table 22. Two sets of accuracy parameter values for each study were required for the model, one for predicting delivery within 7 days and another for delivery before 37 weeks' gestation. We could not obtain 37 week data for one of the studies involving the comparison of fFN with Actim Partus (Bruijn et al. 2016).[45, 46] For this study, therefore, we imputed specificity at 37 weeks from the corresponding 7-day specificity rate so as to obtain a 37-week false positive rate aligned with the UK study of fFN and Italian study of Actim Partus.[56, 102] Similarly, we imputed 37 week sensitivity values for the analysis of Actim Partus vs. fFN (50 ng/ml threshold) based on our meta-analysis of 7 day accuracy data, using the Italian study data.[56] We varied these values in sensitivity analyses.

Table 22 Diagnostic accuracy values used in the PenTAG model

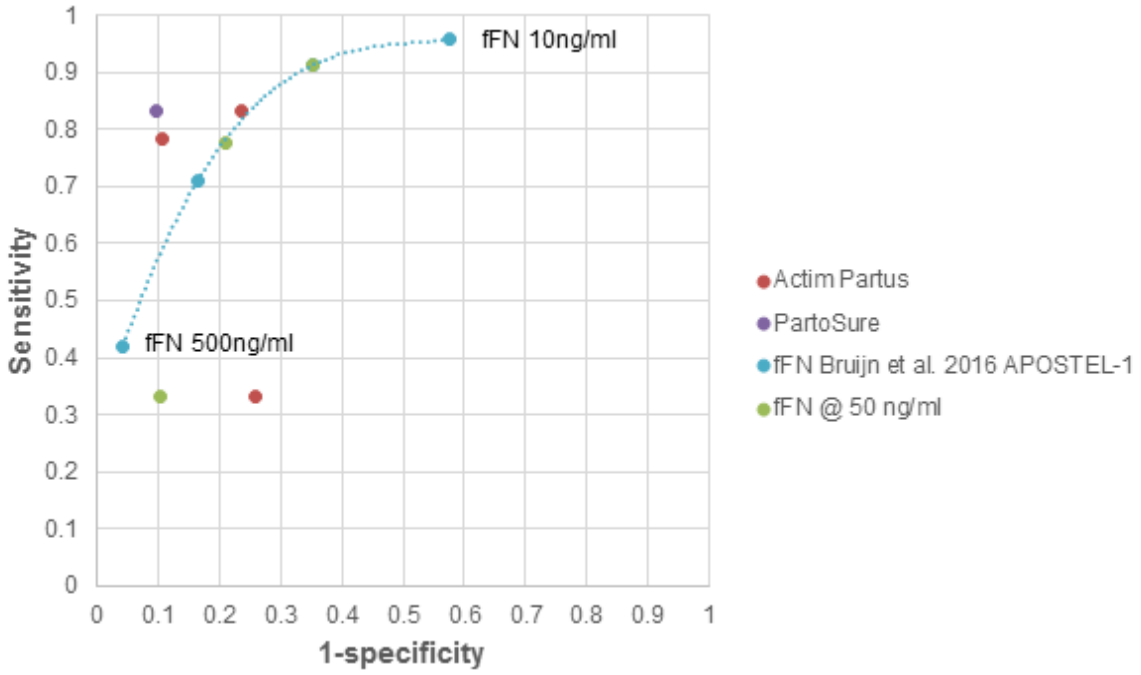
Study	Diagnostic test	N	Sensitivity	Specificity	Probability distribution sensitivity	Probability distribution specificity	Specificity
	Delivery within 7-days						Delivery <37 weeks
Bruijn et al. 2016 APOSTEL-1 [45, 46]	fFN @ 10 ng/ml	350	0.957	0.423	Beta(66,3)	Beta(119,162)	0.458[1]
	fFN @ 50 ng/ml	350	0.913	0.648	Beta(63,6)	Beta(182,99)	0.686[2]
	fFN @ 200 ng/ml	350	0.710	0.836	Beta(49,20)	Beta(235,46)	0.866[3]
	fFN @ 500 ng/ml	350	0.420	0.957	Beta(29,40)	Beta(269,12)	0.972[4]
	Actim Partus	350	0.783	0.893	Beta(54,15)	Beta(251,30)	0.929[5]
Hadzi-Lega et al. 2017[47]	PartoSure	57	0.833	0.902	Beta(5,1)	Beta(39,12)	0.919[6]
	Actim Partus	57	0.833	0.765	Beta(4,1)	Beta(46,5)	0.764[6]
Cooper et al. 2012[1]	Actim Partus	349	0.333	0.741	Beta(2,4)	Beta(254,89)	0.740
	fFN @ 50 ng/ml	349	0.333	0.898	Beta(2,4)	Beta(256,29)	0.946
Abbott et al. 2013[102]	[redacted]						
Meta-analysis by AG[45, 46, 53, 58, 59]	Actim Partus	963	0.832	0.879	Beta(150,30)	Beta(689,94)	0.920[5]
	fFN @ 50 ng/ml	963	0.683	0.872	Beta(123,57)	Beta(683,100)	0.909[5]

Notes: 1 Assumption: False positive rate (FPR) for delivery before 37 weeks was 6% lower than FPR for delivery within 7 days of testing. 2. Assumption: FPR for delivery before 37 weeks was 11% lower than FPR for delivery within 7 days of testing. 3. Assumption: FPR for delivery before 37 weeks was 18% lower than FPR for delivery within 7 days of testing. 4. Assumption: FPR for delivery before 37 weeks was 35% lower than FPR for delivery within 7 days of testing. 5. 7-day FP rate of test times ratio of false positive rate for delivery less than 37 weeks relative to false positive rate for delivery≤ 7 days of corresponding test in Riboni et al.[56] 6 Imputed: 7-day FP rate of test times ratio of false positive rate for delivery less than 37 weeks relative to false positive rate for delivery≤ 7 days of corresponding test in Cooper et al.[1]

Source: Further details are presented in section 2.2.6.1 and Table 6,

Figure 14 depicts the differences in diagnostic accuracy parameters for predicting delivery within 7 days across individual studies in Table 22 (Abbott et al. 2013) by index test and fFN at the 50 ng/ml threshold. The results from the study by Cooper (2012), which compared Actim Partus with fFN at 50 ng/ml, are the two outlying points at the bottom of the graph; as discussed in Chapter 2 it is unclear whether this study used a laboratory or non-laboratory fFN test.[1] Since none of the new tests appears to be superior in terms of both accuracy measures, adopting any one of them implies a trade-off of specificity against sensitivity relative to fFN at 50 ng/ml. Not reflected in the figure is the extent of sampling uncertainty, which may be exemplified by the case of the data for PartoSure, derived from a single study of 57 subjects.[47]

Figure 14 Empirical summary ROC points across evaluated studies (7 day)



6.1.5.2 Background neonatal risks parameter values

Our model included underlying risks of neonatal mortality and adverse events in terms of respiratory distress syndrome (RDS) and intraventricular haemorrhage (IVH), similar to the model that informed the NICE 2015 guidelines on preterm labour diagnosis and treatment, but with data adjusted for steroid use in routine practice (Table 23 and Figure 15).[26] We use the latest estimates of baseline mortality risks by gestational age from the ONS (ONS 2016). The risks of RDS and IVH were derived from Medscape data compiled by Michael Ross and available at https://emedicine.medscape.com/article/260998-overview#a5 (last accessed 20/11/2017).

The baseline risk values in the model are intended to measure neonatal risks in the absence of ANS treatment. Thus the values in Table 23 have been adjusted to subtract the effect of steroids use in routine practice, using the formula:

$$Baseline\ risk = \frac{Unadjusted\ risk}{1 + P_{ANS} * (RR_{ANS} - 1)}$$

where 'Unadjusted risk' is the risk estimate as reported in the data source, P_{ANS} is the prevalence of ANS use in routine practice, and RR_{ANS} is the relative risk of mothers given ANS relative to those not given ANS, and 'baseline risk' is the adjusted risk estimate for each outcome and gestational age reported in Table 23. The adjustment acknowledges the fact that the observed risk in the national statistics is a weighted average of the risk of those who receive and do not receive ANS, where the weights are given by the proportion of

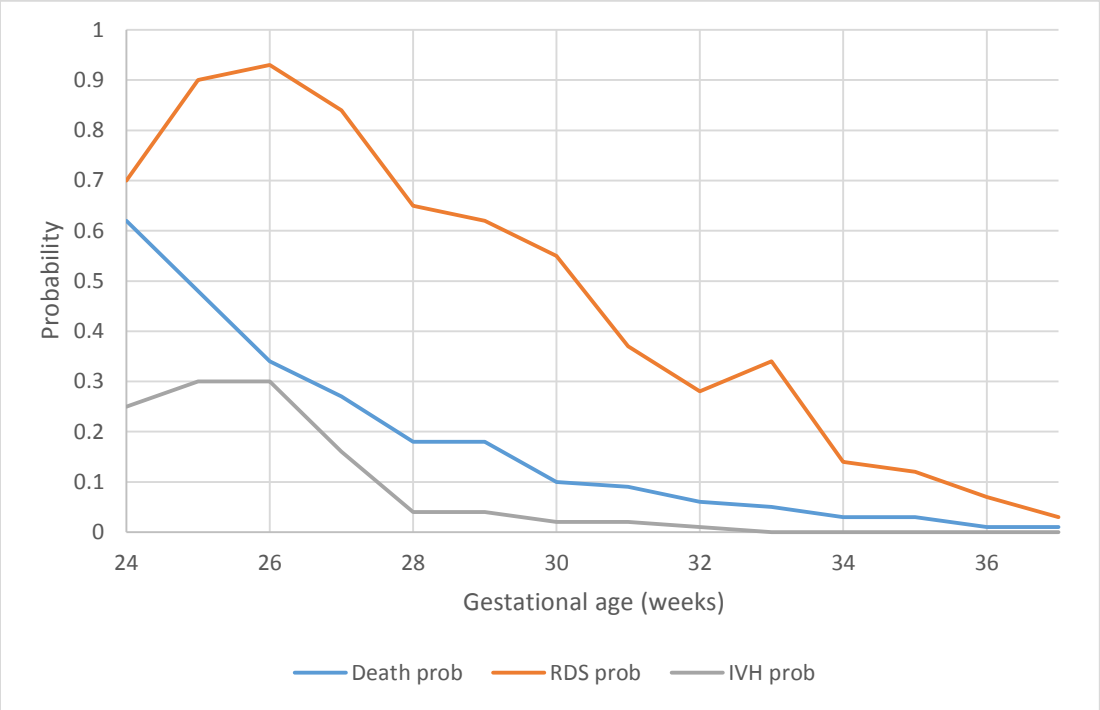
women receiving and not receiving ANS. At lowest extreme, the baseline risk will be equal to the unadjusted risk when no women are treated with ANS or when ANS has no effect on the risk (i.e. RR_{ANS} = 1), and increases with ANS use and the effectiveness of ANS to a maximum of 1/(1- P_{ANS}) times the unadjusted risk (i.e. when RR=0). According to the National Neonatal Audit Programme (NNAP), the most representative data source on ANS use in England, Scotland and Wales, 83% of mothers of babies born between 24 and 34 weeks' gestation in 2013 (the year of our neonatal mortality data) were given at least one dose of ANS (RCPCH 2016).[103] Since NNAP does not produce data by gestational age, we assume such value applies to all gestational ages. We identified one study (Travers et al. 2017)[104] that reports treatment effects of ANS by gestational age (range 23 to 34 weeks) for death before discharge and used that to derive RR_{ANS} (Grant 2014)[105] for neonatal death in Table 23; we adopted the ANS treatments effects on severe intracranial haemorrhage by gestational age from the same source to approximate the RR_{ANS} for IVH. We could not find estimates for ANS treatment effects on RDS by gestational age and thus assumed a constant value of RR_{ANS} for this outcome, from the source described in section 5.1.5.3 (Travers et al. reports treatment effects estimates for bronchopulmonary dysplasia, but the effects were so imprecisely estimated that point estimates implied RR_{ANS}>1 for all but three gestational ages).

Table 23 Baseline risk of neonatal adverse events

	Death		RDS		IVH	
Gestational Age (weeks)	**risk**	**probability distribution Beta(α,β)**	**risk**	**probability distribution Beta(α,β)**	**risk**	**probability distribution Beta (α,β)**
24	0.57	Beta(571,163)	0.70	Beta(408,326)	0.25	Beta(251,483)
25	0.44	Beta(480,244)	0.90	Beta(590,134)	0.30	Beta(309,415)
26	0.32	Beta(403,424)	0.93	Beta(695,132)	0.30	Beta(468,359)
27	0.24	Beta(362,537)	0.84	Beta(650,249)	0.16	Beta(280,619)
28	0.20	Beta(402,731)	0.65	Beta(650,483)	0.04	Beta(71,1062)
29	0.13	Beta(310,975)	0.62	Beta(818,467)	0.04	Beta(78,1207)
30	0.10	Beta (217,1368)	0.55	Beta(808,777)	0.02	Beta(59,1526)
31	0.08	Beta(318,1715)	0.37	Beta(776,1257)	0.02	Beta(60,1973)
32	0.05	Beta(236,2653)	0.28	Beta(771,2118)	0.01	Beta(40,2849)
33	0.04	Beta(320,3738)	0.34	Beta(1414,2644)	0.00	Beta(0,4058)
34	0.03	Beta(429,6368)	0.14	Beta(892,5905)	0.00	Beta(0,6797)
35	0.02	Beta(516,9518)	0.12	Beta(1128,8906)	0.00	Beta(0,10034)
36	0.01	Beta(547,19561)	0.07	Beta(1319,18789)	0.00	Beta(0,20108)
37	0.01	Beta(687,43,773)	0.03	Beta(1458,43002)	0.00	Beta(0,44460)

Key: IVH, Intraventricular haemorrhage; RDH respiratory distress syndrome
Source: UK stillbirth and neonatal mortality rates (ONS 2016) and US data from Medscape (https://emedicine.medscape.com/article/260998-overview#a5, last accessed 20/11/2017). Adjusted used data on ANS treatment effects from Travers et al. 2017 and steroid use in routine practice from the Neonatal National Audit Programme (RCOCH 2016).

Figure 15 Baseline risks of neonatal adverse events



Source: UK stillbirth and neonatal mortality rates (ONS 2016) and rates of RDS and IVH in US data from Medscape (https://emedicine.medscape.com/article/260998-overview#a5, last accessed 20/11/2017).

Since some IVH and RDS cases result in fatality, to calculate the number of infants who live to adulthood with these conditions we follow the NICE 2015 Guideline model and multiply the incidence of RDS and IVH by 1 minus the probability of neonatal mortality among neonates with these events, which is assumed constant across gestational ages.[26] We searched the literature and identified a new source of data on the probability of death related to IVH, which we adopted in our model.[106] We found no new data on the probability of neonatal death among RDS cases and thus used the value from the NICE guidelines, which was obtained from US data for 2004 (CDC 2007).[26] The values are 0.054 (Beta[875,15393] in probabilistic sensitivity analyses) for RDS and 0.205 (Beta[76,394]) for IVH.

6.1.5.3 Steroid treatment

In the model, women who test positive are treated with ANS. In the scenarios where women present at level 1 or 2 maternity hospitals at less than 28 weeks' gestation, they also receive tocolysis and are transferred in utero to a level 3 unit. The model does not account for any possible effects of tocolysis in terms of delaying preterm delivery.

For our base case analysis we used treatment effects parameter values for ANS administration from results reported by the Effective Perinatal Intensive Care in Europe (EPICE) study, a prospective cohort study that collected data from 19 regions in 11 European countries in 2011 and 2012.[36] This study was selected as the largest, and most

representative source of data on ANS effectiveness in reducing neonatal mortality and morbidity of very preterm infants by time to delivery. The EPICE study produced an analysis of the association of administration-to-birth intervals with morbidity and mortality in 4,594 infants born at gestational ages between 24 and 31 weeks. Given its large sample size, the study was able to analyse the outcomes associated with corticosteroids given a few hours before birth relative to outcomes at longer administration-to-birth intervals. The study concluded that the ANS may be effective even when administered up to 3 hours before delivery, which was expected to reduce mortality relative to no ANS by 26%.[36] The authors reported that 77.9% of the 1,111 women who received ANS less than 24 hours before delivery received only one dose of ANS. Treatment effects on IVH were also derived from this source. For treatment effects on RDS, we used data from the Cochrane Systematic Review of RCT on the effectiveness of ANS relative to no treatment or placebo, which was also the source of values for sensitivity analyses. The main findings from the Cochrane review are summarised in Table 3. Since subgroup analysis produced no evidence that rupture of membrane status led to different rates of neonatal death, fetal death, RDS, IVH or birthweight in infants exposed to corticosteroids, we decided to use the overall treatment effect estimates in our model.[37]

Table 24 Treatment effect of antenatal corticosteroids

	Relative Risk (95% CI)	Source
Fetal mortality	0.98 (0.74 to 1.30)	Meta-analysis; participants = 6729; studies = 15
Neonatal mortality	0.69 (0.59 to 0.81)	Meta-analysis; participants = 7188; studies = 22
RDS	0.66 (0.56 to 0.77)	Meta-analysis; participants = 7764; studies = 28
Moderate to severe RDS	0.59 (0.38 to 0.91)	Meta-analysis; participants = 1686; studies = 6
IVH	0.55 (0.40 to 0.76)	Meta-analysis; participants = 6093; studies = 16
Severe (Grades 3 and 4) IVH	0.26 (0.11 to 0.60)	Meta-analysis; participants = 3438; studies = 6
Chronic lung disease	0.86 (0.42 to 1.79)	Meta-analysis; participants = 818; studies = 6

Key: IVH, intraventricular haemorrhage; RDS, respiratory distress syndrome
Source: Cochrane review (Roberts et al. 2017)[37]

The effectiveness of steroids depends on the time from ANS administration to delivery (Table 25) displays the treatment effect model parameter values for the base case and sensitivity analysis, which reflect the reduced effects of ANS when given earlier than 7 days before birth.

Table 25 Estimated effectiveness values of steroid treatment used in the model

Parameter	Base case value	Scenario analyses	Probabilistic distribution for sensitivity analysis [lognormal(mean, SD)]	Source
Treatment effects: Neonatal mortality RR (95% CI)				
ANS≤ 7 days vs. no ANS	0.5 (0.4-0.6)	0.69 (0.59 to 0.81)	Log normal (0.50, 0.093)	Base case: ANS 24h-7 days, adjusted estimate Table 2 in Norman et al. 2017.[36] Scenario: Cochrane review (did not distinguish by timing of ANS)
ANS>7 days vs. no ANS	1	0.7 (0.6-0.9) 0.69 (0.59 to 0.81)	Not varied: fixed at 1	Scenario: Cochrane review (Roberts et al. 2017)[37]; Norman et al. 2017.[36]
Treatment effects: RDS RR (95% CI)				
ANS≤ 7 days vs. no ANS	1	0.66 (0.56 to 0.77)	Log normal (0.66, 0.079)	Base case: Cochrane review 2017[37]
ANS>7 days vs. no ANS	1	0.66 (0.56 to 0.77)	Not varied: fixed at 1	Scenario Analysis: Cochrane review 2017[37]
Treatment effects: IVH RR (95% CI)				
ANS ≤7 days vs. no ANS	0.6 (0.5-0.9)	0.55 (0.40 to 0.76)	Log normal (0.60, 0.207)	Base case: ANS 24h-7 days, adjusted estimate Table 2 in Norman et al. 2017.[36] Scenario: Cochrane review 2017.[37]
ANS>7 days vs. no ANS	1	0.55 (0.40 to 0.76) 0.8 (0.6-1.2)	Not varied: fixed at 1.	Scenario: adjusted estimate Table 2 in Norman et al. 2017;[36] Cochrane review 2017.[37]
Treatment effects: Birthweight mean difference, grams (95% CI)				
ANS≤7 days vs. no ANS	*0*	*0*	Not varied: fixed at 0	Assumption based on Roberts and Dalziel 2006.[35]
ANS>7 days vs. no ANS	*0*	*-147.0 (-292.0, -2.0)*	Not varied: fixed at 0	Base case: Assumption based on low quality of evidence in Roberts and Dalziel 2006;[35] Scenario: WHO 2015[107], Roberts and Dalziel 2006.[35]

Key: ANS antenatal steroids; RR, relative risk

6.1.5.4 Health-related quality of life

We conducted a systematic search of the literature for utility values of neonatal outcomes in the model: mortality, respiratory distress syndrome and intraventricular haemorrhage. The details of the search strategy, identification and data extraction from the identified studies are provided in Appendix 1. In this section we summarise our findings.

6.1.5.4.1 Summary of identified studies

A total of 28 studies were identified from screening full-texts as containing information useful for obtaining or deriving utility parameters for the model, given the populations studied (i.e. either preterm children, or mothers). These studies are broadly summarised in Appendix 3.

Of these 28 studies, 24 assess the outcomes of children born preterm. The details of these studies are summarised in Table 60. Nine additional papers were cited as sources for utilities in some of these studies. Parameter values from these additional papers are presented in Table 61. The remaining four papers assess the outcomes of mothers. These studies are summarised in Table 62.

6.1.5.4.2 SF-36 mapping and extraction of utilities

None of the studies that were found directly measured utilities based on the EQ-5D. However, various mapping functions exist which allow 36-Item Short Form Survey (SF-36)[108] summary measures to be converted into EQ-5D utilities.[109] We make use of a mapping function obtained from Rowen et al.[110] in order to perform this conversion, as it was deemed the most appropriate study, based on regression variables and the population sample used. A more detailed discussion of mapping studies can be found in Appendix 3.

6.1.5.4.3 Relevant studies for utilities of IVH, of RDS and for mothers

Only one paper considers the quality of life for preterm children with IVH, separated into two severity groups: level 0-2 IVH with no periventricular leukomalacia (PVL); and level 2-4 IVH with/without PVL.[111] However, a suitable mapping to EQ-5D utility for the health-related quality of life measure they used, developed by the Centers for Disease Control and Prevention (CDC), could not be found.

Likewise, only one paper considers quality of life for preterm children with RDS. This study measures SF-36 scores, but does not report them.[112]. We were unable to obtain the SF-36 data after contacting the corresponding author.

The quality of evidence on the quality of life of mothers of preterm children is sparse. Only two studies consider mothers of preterm children specifically.[113, 114] The first is an abstract that reports only physical and mental health SF-36 mean summary scores, while the second reported MAPP-QOL scores. Neither of these could be reliably mapped to EQ-5D utilities.

Couto et al. assess the quality of life in mothers in Brazil who have had at least one of four previous adverse pregnancy outcomes.[115] While preterm birth is one of the four outcomes that is an inclusion criterion (along with early neonatal death, recurrent abortion, and fetal death), we are not provided with separate utilities for each outcome individually.

A more detailed discussion of all reviewed studies for the utilities of preterm survivors; IVH; RDS; and mothers, is provided in Appendix 3.

6.1.5.4.4 Utilities for reduced birthweight

In order to assess whether there is any quality of life impact from reduced birthweight as a result of not receiving treatment, papers that were identified from title and abstract screening were searched to find studies that contained both sufficient birthweight and utility data. One study was identified, and the author was able to provide the raw data upon request.[116]

A number of regression specifications were estimated using random effects estimators, in order to find the effects of birthweight on utility (see A4.9). However, the coefficient estimates for birthweight and squared birthweight were not statistically significant at the 5% level in any model. Furthermore, the simplest specification (including only birthweight and squared birthweight) failed the Likelihood Ratio test when compared to a specification that included gestational age, sex, and time dummies. Therefore, the analysis found insufficient evidence of a birthweight effect on utility, and so it is assumed that there is no utility loss from reduced birthweight alone. Further details of the statistical analysis can be found in A4.9.

6.1.5.4.5 Utility parameters selected for the economic model

Based on the discussion in Appendix 3 utilities selected as the most appropriate base case values for the economic model are reported in Table 26. Proxy utilities for RDS and IVH were obtained from Carroll and Downs, as the source for the study by Bastek et al.[117, 118]

In practice, not all children with RDS go on to develop severe persistent asthma (the proxy for RDS used in the model). Based on feedback from a neonatologist, we apply this proxy utility to 56% of all RDS cases. This figure is from a UK based study that found 56% of children born extremely preterm had abnormal baseline spirometry at age 11.[119] The remaining 44% of RDS cases are assumed to incur no additional QALY loss, relative to a preterm survivor.

For IVH, the proxy used (moderate cerebral palsy) is too severe for infants with IVH grades below III. Upon consultation with a clinical expert, long term outcomes from IVH grades below III are thought not to differ greatly from those of preterm survivors in general. Therefore, we only apply the utility of moderate cerebral palsy to incidences of IVH that are at grades III or IV.

The utility for preterm survivors was obtained using mapped SF-36 scores.[110] The utility computed from the UK study was used for the base case, while the minimum and maximum utilities in the remaining papers were selected to provide a range.[120-122]. This follows the principles outlined in the NICE technical support document.[123]

Infant mortality is assumed to have a utility of 0, as in the 2015 NICE guidelines model.

We also considered measuring the health-related quality of life outcomes of the mother. We identified published evidence on longer-term utility for mothers with previous adverse pregnancy or neonatal outcomes from Couto et al.[115] This can be used as a proxy for mothers who have preterm children that suffer an adverse outcome, with the caveat that it will likely be an overestimate for infant mortality, and an underestimate for IVH or RDS. The lower (upper) bound for the two utilities imputed from Couto et al. is calculated by taking the lower (upper) value of the 95% confidence interval for each of the eight SF-36 dimensions, and generating a mapped EQ-5D from this vector. This provides a relatively pessimistic (wide) estimate of the range of utilities, which is desirable given the caveat attached to the adverse outcome utility value.

Table 26 Utilities selected for the economic model

Variable	For	Source	Measure	Utility	Range
'Severe' RDS (severe persistent asthma used as proxy)	Child	Carroll and Downs [117]	TTO	0.85	0.84-0.86†
IVH Grades III-IV (moderate cerebral palsy used as proxy)	Child	Carroll and Downs via Bastek et al.[117, 118]	TTO	0.76	0.66 – 0.84*
Death	Child	Assumption. Upper bound from Vandenbussche et al.[124]	SG (upper bound only)	0	0 – 0.02
Preterm survivor	Child	Cooke [120]	SF-36†	0.879	0.846 – 0.901ω
Mother with previous adverse child outcome	Mother	Couto et al.[115]	SF-36†	0.644	0.556 – 0.652π
Mother with no adverse child outcome	Mother	Couto et al.[115]	SF-36†	0.834	0.768 – 0.843π

Notes: † Range calculated as a 95% confidence interval, based on the data from Carroll and Downs.[117] * Range taken directly from Bastek et al.[118]. These represent the minimum and maximum values found in their literature search for these utilities.; ω Range taken from two of the five studies reporting SF-36 scores (the minimum and maximum utilities reported amongst the five studies).[121, 122] ;† SF-36 means for the eight dimensions were mapped onto EQ-5D utilities using a quadratic model.[110]; π Range is generated by generating a 95% confidence interval for the eight SF-36 means, and mapping all lower bounds and all upper bounds to EQ-5D using a linear model.[110] These represent a wider estimate of the range than if a 95% confidence interval was provided for the EQ-5D measure directly.

Key: IVH, intraventricular haemorrhage; RDS, respiratory distress syndrome; TTO time trade off

6.1.5.4.6 QALY calculations and comparison with parameters used in NICE guidelines model

Preterm survivors

NICE Guidelines

The model used for tocolytic treatment in the 2015 NICE guidelines for preterm labour provides utilities for preterm survivors through the gestational age range from 24 to 34 weeks.[26] The assumption made was that those surviving to age 1 would live 80 years (based

on life expectancy in 2015 in England and Wales) at a utility value of 0.82 per year (based on a UK population norm).[2] This was then discounted by the standard rate of 3.5%, and multiplied by the probability of survival at each gestational age. Therefore, the utility of preterm survivors in the guideline model was not based upon data specifically from individuals born preterm.

PenTAG model

In contrast, the utility selected for our model is based on the mapped SF-36 score of preterm survivors assessed between ages 19 and 22.[120] The range is generated by taking the minimum and maximum mapped SF-36 values out of four other SF-36 follow-up studies in preterm individuals, assessed at a similar point in the life cycle (from 20 to 31 years of age).[121, 122, 125, 126]

Rather than assuming life expectancy, our model uses ONS 2014-16 life tables for survival proportions at each age. These are multiplied by baseline population utilities for each age, which are derived from a regression equation that was fitted to EQ-5D data from the Health Survey for England by Ara and Brazier.[127] These baseline utilities account for the natural decline in health related quality of life with ageing. Some extrapolation was necessary at the beginning and end of the life horizon, as the life tables cover the age range 0-100, whilst the Ara and Brazier regression equation was obtained by fitting to data with an age range of 16-98. Finally, the utilities for each health state from Table 26 are multiplied to these baseline utilities at each age.

The probabilities of survival after 1 year at each gestational age were obtained from 2013 ONS data for England and Wales. These were applied to the resulting overall discounted sum of utilities over the lifetime to obtain total QALYs. Our method improves upon that of the 2015 NICE guidelines model by taking population aging and survival into account, whilst preserving NICE's use of survival probabilities by gestational age. Alternative QALY values were also calculated by excluding the population aging effect, and assuming a fixed life expectancy, in order to provide QALYs that use the same methodology as the NICE guidelines, but with updated health state utilities.

We were unable to find preterm-specific life tables or survival curves across the lifetime. Two studies were found using data from Sweden, which reported hazard ratios for preterm children.[128, 129] Both of these studies showed a very small increase in mortality until 5 years of age, and negligible increases in mortality beyond 5 years. Because of this, we did not apply additional adjustments to the survival probabilities from preterm birth after 1 year.

[2] The NICE guidelines state that this norm was established in either 1982 or 1983 (two different years are stated at different places in the evidence report[26]) Therefore, it is somewhat outdated.

Table 27 summarises the base case total QALYs for preterm survivors by gestational age. The NICE guidelines model stops at 34 weeks gestational age, and therefore no NICE QALYs are reported for the 35 week and 36 week gestational ages.

Table 27 Total discounted QALYs for preterm survivors, by gestational age

Gestational Age (weeks)	P (survive 1st year) from ONS 2013 data[130]	Weighted discounted QALY using age and survival adjustments	Weighted discounted QALY using NICE method (=Max QALY x P(survive 1st year))	Weighted discounted QALY used in 2015 NICE guidelines
24	0.5934	13.23	14.44	19.92
25	0.7621	16.99	18.55	20.89
26	0.8289	18.48	20.17	21.27
27	0.8917	19.88	21.70	21.69
28	0.9120	20.33	22.19	22.18
29	0.9544	21.27	23.23	22.44
30	0.9679	21.57	23.55	22.61
31	0.9733	21.69	23.69	22.52
32	0.9833	21.92	23.93	22.53
33	0.9870	22.00	24.02	22.58
34	0.9904	22.08	24.10	22.61
35	0.9916	22.10	24.13	N/A
36	0.9955	22.19	24.23	N/A

Key: ONS, Office of National Statistics; QALY, quality adjusted life year

IVH and RDS

The 2015 NICE guidelines model for tocolytic treatment assumes a 'dummy value' for the lifetime QALY loss from RDS, due to its variable prognosis. This value was chosen to be marginally lower than the lifetime QALY loss from IVH. The economic model in Bastek et al.[118] uses a 10 day ICU stay as a proxy for RDS, but also provides a proxy for chronic respiratory disease as being the utility of moderate persistent asthma. The original time trade-off values for these parameters from Carroll and Downs are identical at 0.91.[117] However, this value is higher than the mapped utility of preterm survivors found in the literature. Therefore, we opted to use the utility of severe persistent asthma from Carroll and Downs as a proxy for the lifetime effects of RDS (0.85).

The QALY loss from IVH of 4.5 in the Guidelines model was based on the value for Intracranial Haemorrhage (ICH), which was assumed to be 1/3 of the QALY loss from moderate to severe cerebral palsy.[94] The source for the health state utility upon which this QALY loss was derived appears to have been obtained from Pham and Crowther (2003) as the utility of permanent neurological sequelae, assessed by antenatal or emergency midwives.[131] However, a very small sample size of 14 was used to elicit this result, via the standard gamble method.

In comparison, the proxy of moderate cerebral palsy used in Bastek et al. for IVH was obtained from Carroll and Downs.[117, 118] This appears to be a more reliable source, given that the TTO elicitation task was performed on a much larger (and more relevant) sample of 4,016 parents.

The same age and survival adjustment methodology was used to compute the discounted total QALYs over the lifetime for children with RDS and IVH as with preterm survivors. Table 28 summarises the total QALY loss from RDS and IVH (relative to a preterm survivor). It should be noted that these values imply the assumption of no effect of IVH and RDS on survival after the first year.

Table 28 Total discounted QALY loss for RDS and IVH as used in the economic model

Outcome	PenTAG base case	NICE guidelines value
RDS (applied to all cases)	0.74	3.85
RDS (applied to 56% of cases)	0.41	2.16 (inferred)
IVH (applied to all grades)	3.02	4.50
IVH (applied to grades III and IV – 30% of cases)	0.91	1.35 (inferred)

Key: IVH, Intraventricular haemorrhage; QALY, quality adjusted life year; RDS, Respiratory distress syndrome

Mortality

Whilst the utility for infant mortality is assumed as 0 (as in the NICE Guidelines model), an upper bound of 0.02 was also included for the purposes of scenario analysis.[124] This allows for a scenario in which a child is deemed to gain some utility for the short period of life before they die.

Mothers

In common with all other previous models of preterm labour, the NICE Guidelines model does not account for the health-related quality of life of mothers of preterm children after adverse outcomes. While data are extremely limited, we use the utility mapped from the SF-36 scores reported in Couto et al. of mothers that have had previous adverse pregnancy outcomes as a proxy for the utility for a mother after infant mortality in the model to conduct exploratory scenario analyses.[115] We consider two scenarios. First, the mother suffers the adverse pregnancy outcome utility for her remaining lifetime. Second, the mother suffers the adverse pregnancy outcome utility for 10 years, and then reverts to the utility for mothers with no previous adverse pregnancy outcomes. ONS 2013 data states that the mean age of mothers at birth is 30, and this is echoed by findings from the Assessment of Test Accuracy section earlier in this report. This is used as the starting age from which to compute QALYs for mothers.

It has been previously discussed that this utility is too broad to capture a mother's utility for each individual child outcome accurately. However, it does allow us to contrast the potential average disutility for a mother who loses her child.

Table 29 Total discounted QALYs for mothers (assuming age at birth of 30)

	Lifetime QALYs
Child dies (applied for lifetime)	13.45
Child dies (applied for 10 years following birth	15.94
Child survives	17.42

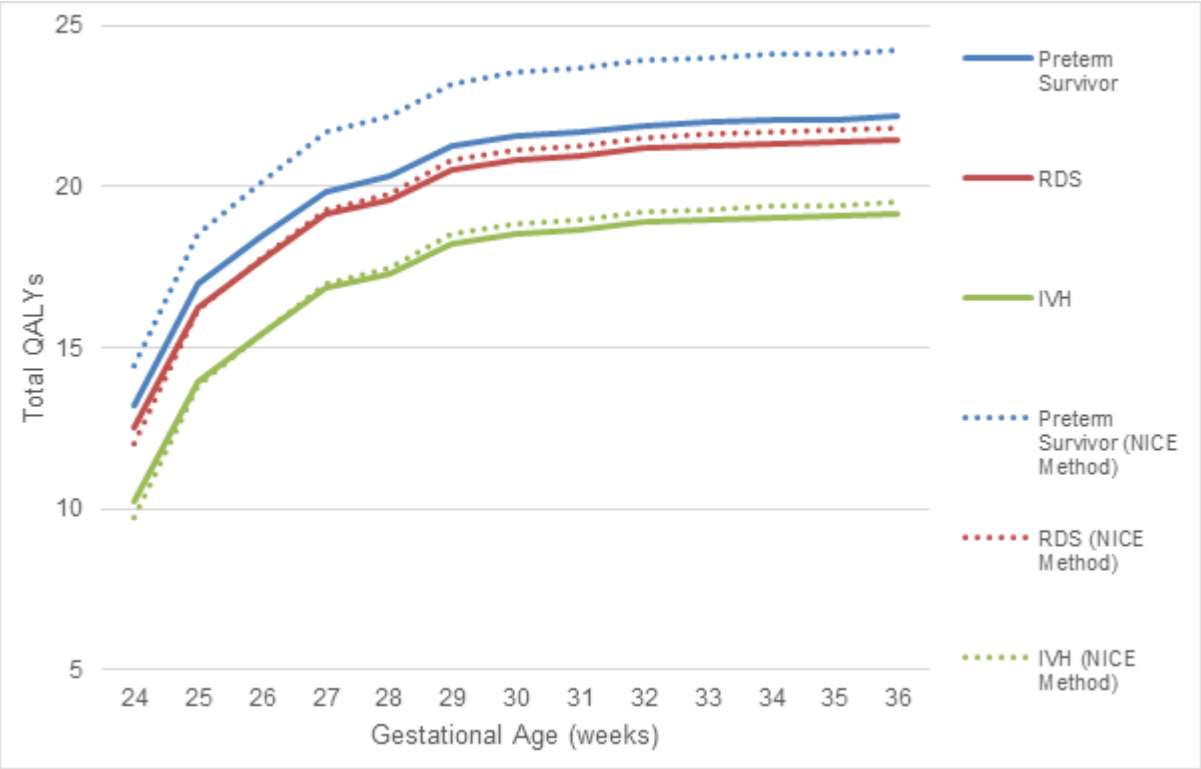
Key: IVH, Intraventricular haemorrhage; QALY, quality adjusted life year;

6.1.5.4.7 Summary of all child-related QALY values used in the model

The base case total discounted lifetime QALYs for children are summarised in Figure 16. For comparison, total discounted lifetime QALYs calculated using the NICE guidelines method (i.e. without age and survival adjustments across the lifespan) are denoted by dotted lines in the same figure.. As can be seen by comparing these graphs, the NICE guidelines method for obtaining QALYs appears to:

- Overestimate total QALYs for preterm survivors
- Overestimate the QALY losses from RDS and IVH

Figure 16 Total lifetime QALYs for preterm children by gestational age and health state, using age and survival adjustment (solid lines), and using NICE's unadjusted method (dotted lines)



Key: IVH, Intraventricular haemorrhage; QALY, quality adjusted life year; RDS, respiratory distress syndrome
Notes: QALY losses in the graph are not scaled to apply to only severe cases of RDS and IVH. This is to enable clearer visual comparison between the two QALY calculation methods.

6.1.5.5 Costs

The cost parameter values used in the model are presented in Table 30. The costs of the three tests include the time involved in the costs of acquiring the test itself and the time required for a midwife to apply the test. The costs of ANS injection and tocolysis treatment with Atosiban, as well as hospital admission and in-utero transfer were obtained from a published UK costing study conducted during 2009/2010 in a London university hospital (Parisaei 2016).[132] These costs were inflated to 2016 prices using the Health care and prices index in the Personal Social Services Research Unit (PSSRU) publication of unit costs of health and social care.[133] The cost of in-utero transfer in this study only include the costs of the ambulance transfer service; to this, we added the cost of arranging a transfer service and the cost of a midwife accompanying the baby during the transfer.

The costs of adverse neonatal outcomes were derived from our analysis of the National Neonatal Research Database, which contains selected data from the Badger.net neonatal electronic health records for the years 2014/2015.[134] In these data, the number of days spent at BAPM levels I-IV were applied to the respective HRG tariffs for 2016. We used HRG reimbursement tariffs as opposed to the HRG reference costs on the advice that the latter are unlikely to reflect actual resource use at the different levels of neonatal care, given the

common accounting practice of arbitrarily apportioning costs to the different levels of care by hospitals in their HRG reference cost reports (Eleri Adams personal communication February 2017). It was therefore thought that HRG tariffs would better reflect the resource use in neonatal units, at least until the new HRG reference costs for neonatal critical care become available.[135]

Unlike the economic analysis that informed the NICE 2015 Guidelines on diagnosis and treatment of preterm labour, our model accounts for the additional costs of saving a preterm neonatal life. There are two offsetting effects on costs from saving the life of an infant. The cost per inpatient hospital day of an infant who dies before discharge is likely to be greater than for a surviving infant. However, the length of hospital stay of the surviving infant is much larger than for an infant that dies. Overall, the cost of the length of hospital stay dominates and we estimate in national data (Badger)[49] that saving a baby by means of timely ANS treatment has the knock-on consequence of increasing neonatal hospital costs to the NHS by £22,834. It is noteworthy that in a modelling study of the public sector costs of a preterm survivor up to age 11, it is estimated that neonatal hospital costs account for ~90% of the total.[136]

The costs of RDS and IVH were estimated from the Badger dataset. In the dataset available to us, we did not have the information required to identify cases of IVH or RDS. We therefore estimated the difference in cost between those receiving and those not receiving neonatal intensive care (BAPM level I), after adjusting for gestational age, birthweight, sex, multiparous pregnancy, type of labour (spontaneous versus induced) and mode of delivery (vaginal vs. caesarean), and assumed that the resulting estimate was approximately equal to the cost of IVH or RDS in the model. We recognise that BAPM level I care results also from other causes such as sepsis, necrotising enterocolitis, pulmonary haemorrhages, and may be due to reasons as varied as nutrition, surgery, chest drains or congenital abnormalities. However, some of these other causes may be affected by antenatal steroids but not amenable to be formally accounted in our model due to lack of evidence. To that extent any bias that may result from this assumption may be limited by the extent to which it compensates for those other unmeasured benefits of ANS in our model.

Table 30 Cost parameter values (2016 prices)

Cost parameter	Unit cost (£)	Definition, price year and source
qfFN test	65	Based on 15 minutes of midwife time, with unit costs from PSSRU (Curtis and Burns 2016)[133] Costs of the test excluding VAT (typically £35). Hologic 2017 Request for Information from NICE. Due to lack of data this figure does not include the cost of test failures.
pIGFBP-1	35	Based on 10 minutes of midwife time, with unit costs from PSSRU (Curtis and Burns 2016)[133] and £15 cost of Medix test exc. VAT; Alere 2017 Request for information from NICE.
PAMG-1	52	Based on 10 minutes of midwife time, with unit costs from PSSRU (Curtis and Burns 2016).[133] Cost of the test in the UK excluding VAT. Parsagen 2017 Request for Information from NICE.
Maternal steroid injection	5	UCLH 2012; Parisaei 2016[132]
Atosiban+ atosiban infusion equipment	362	Atosiban infusion equipment includes syringe pump, syringe and giving set; Parisaei et al. 2016[132]Dosage or units of doses not given, nor does it include costs of time to administer. Alternative values: BNF 2016, cost of solution for infusion Atosiban acetate 37.5 mg/5ml concentrate for solution for infusion vials, 1 vial £52.82 (hospital only) at maximum or alternatively half the maximum adult dose in BNF 70 of 330.75 mg over 48 hours plus equipment cost in Parise et al. 2016[132].
Inpatient hospital	1,325	Median length of hospital stay, 2 days, times cost of 24-h admission to hospital; Parisaei 2016[132]; Primary Care Trust in London
In utero transfer	965	London Ambulance Service, 2012; Parisaei 2016.[132] It includes 6 hours (Gale et al. 2012)[137] of a modern matron's time to arrange transfer; i.e. 6 times £62 (Curtis and Burns 2016).[133]
Long term healthcare costs of IVH	114,648	Downstream healthcare costs NICE Guideline 2015. These were assumed to be equal to the cost of ICH, and that Grade III and Grade IV ICH equals the cost of cerebral palsy. The calculation used by NICE 2015 seems to be wrong; we assume the correct number is equal to at least £79,000, and use this number when all grades of IVH are considered (alternatively the assumption in the NICE 2015 Guideline model that this reflects 30% of the value for Grade III-IV is used when considering severe IVH only). The value assumes a life expectancy of 60 years and a discount rate of 5% - this was adjusted using ONS 2014-16 life tables and for a discount rate of 3.5%.
Neonatal hospital costs of preterm survivors discharged home/to ward	32,435	PenTAG analysis of Badger data for infants born at gestational ages below 36 weeks in England and Wales in 2013-2014 (n=22,936). Includes the costs of BAPM levels 1,2,3,4,5 (XA01Z, XA02Z, XA03Z, XA04Z, XA05Z) at 2014/15 NHS tariffs. Mean overall length of stay (superspell) of 46 days (potential outcome without death).
Neonatal hospital costs RDS	5,587	OLS adjusted difference in neonatal hospital costs between infants with and without days spent in BAPM level of care 1 in Badger 2014/2015; valued at the national tariffs for BAPM levels 1,2,3,4,5 (XA01Z, XA02Z, XA03Z, XA04Z, XA05Z) in 2014/15 prices. Alternative value: Downstream healthcare costs; NHS Reference Costs 2011/12, XB01Z Paediatric Critical Care, Intensive Care, ECMO/ECLS; NICE Guideline 2015. Alternative value 2:Landry 2012[138], preterm hospitalization cost, 2008 Canadian dollars, cost of medical & pharmaceutical services also given. Adjusted using the HCHS indices and the purchasing power parities to reflect the equivalent costs in the UK in pounds sterling in 2016. The HCHS index for 1990/1991 was calculated by taking the geometric average yearly increase between 1988 and 2006.
Additional neonatal hospital costs: infant dies before discharge	-22,834	Base case: neonatal hospital cost that would have been incurred by a neonatal fatality had preterm child survived; calculated by AG from Badger data. Alternative value: assumption of no costs, as in NICE 2015 Guideline model. Alternative value 2: from Khan et al. 2015.[139]

Note: The costs were adjusted using the HCHS indices (PSSRU; Curtis and Burns 2016)[133] to reflect the equivalent prices in 2016.

6.1.6 Base case analyses

In the base case analysis we consider the case of women presenting to a level 2 care hospital. This is based on level 2 hospitals having the highest frequency of cases, given the mean gestational age at presentation of 30 weeks. In addition we set the prevalence rate of preterm birth within 7 days of testing at 3.0% and the prevalence rate of preterm birth at <37 weeks' gestation of 12.1 %.[102] In order to derive the costs and QALYs of PartoSure for comparison against fFN 50 ng/ml (and other options in the full incremental analyses) we performed an indirect comparison whereby the incremental costs and QALYs of PartoSure relative to fFN 50 ng/ml based on the test accuracy data from Hadzi-Lega et al. 2016 were added to the costs and QALYs of Actim Partus based on the APOSTEL-1 data (Bruijn et al. 2016). Table 31 summarises the main model assumptions, described above.

Table 31 PenTAG model specifications and assumptions

	Base case model specification/assumption	**Comment**
Patient population	Symptomatic women with intact membranes presenting to level 2 hospital, who have not been ruled for preterm labour after clinical examination	Clinical examination is not a relevant comparator for this population, since it precedes the starting point of the analysis Scenario analyses consider women presenting at level 1 and 2 hospitals.
Time horizon	Lifetime	Scenario analyses limit the horizon to delivery and alternatively to neonatal hospital discharge
Diagnostic test protocol/guideline	Complete adherence of treatment decisions to results of diagnostic test	
Differences in clinical outcomes between test options are the result of differences in test sensitivity	Differences in true positive rates result in differences in neonatal mortality and neonatal morbidity outcome (and costs) through the timely use (within 7 days of delivery) of antenatal steroids (ANS). Maternity costs are dependent on test sensitivity. i.e. false negatives (i.e. delivering within 7 days after negative test result) are 'missed' (do not receive ANS).	We vary this assumption in the scenario analysis that limits the analytical horizon to delivery, so that neonatal outcomes and costs are assumed to be the same across strategies and the only difference is in terms of maternity costs, i.e. differences only depend on test specificity.
Adverse events included neonatal mortality and morbidity	Outcomes considered are neonatal death, respiratory distress syndrome, and intraventricular haemorrhage	This is in line with previous model informing the NICE Guidelines on preterm labour diagnosis and treatment (NICE 2015)
Neonatal mortality results in net savings to the NHS	Saving a neonatal life through accurate diagnosis and timely ANS treatment, has the consequence of increasing NHS costs, since the infant saved stays longer in neonatal hospital (although at a slightly less intensive average level of care per day).	We explore the impact on results of assuming that saving a child does not incur additional costs, in scenario analyses
ANS are only effective if given within 7 days of delivery	Infants born more than 7 days after testing positive do not benefit from ANS	In scenario analysis we allow for partial benefit from ANS for those testing positive and given ANS

		earlier than 7 days before preterm birth (i.e. at less than 37 weeks).
In utero transfers are only required for women presenting to a level 1/ 2 hospital at less than 28 weeks' gestation	Transfer to a tertiary hospital	This is in line with NICE guidelines.
Tocolysis	Only used for in-utero transfers; no consequences on clinical effects, only on costs	This is intended to reflect emerging consensus about the benefit/risks profile of tocolysis
Long term costs	Only included those associated with IVH	In line with the model informing the NICE 2015 guideline of preterm labour diagnosis and treatment
Long term quality of life	Assumed those who survive beyond 1 year of life to achieve the average long term quality of life in the general population, regardless of preterm birth status.	In line with the model informing the NICE 2015 guideline of preterm labour diagnosis and treatment; plausibility supported with clinical experts' opinion

Key: ANS, antenatal steroids

6.1.7 Scenario Analyses

We explore the following scenario analyses:

- Alternative study sources of test accuracy data
- Women presenting at tertiary level unit hospitals
- Limiting costs to the neonatal phase
- Limiting costs to the diagnostic phase (until delivery)
- Assuming ANS have (partial) benefits when administered earlier than 7 days before preterm delivery
- Excluding the neonatal hospital costs of infant death
- Include mother QALYs

6.1.8 Probabilistic Sensitivity Analyses

We present probabilistic analyses using information on sampling uncertainty for test accuracy, costs and utilities presented in Table 22, Table 23, and Table 25.

6.2 Results

Since our results vary by gestational age, we present details for the base case of a symptomatic woman presenting at 30 weeks of gestation (the average age on diagnostic accuracy studies), and general results for older, 33 weeks, and younger, 26 gestational ages.

6.2.1 Base case results

The base case deterministic results are presented in Table 32. These are based on the preferred comparative studies APOSTEL-1 and Hadzi-Lega.[45-47] The base case considers

women presenting at 30 weeks' gestation to a level 2 hospital. Whilst all ICERs are positive, they should be interpreted with caution since, other than 'treat all' and fFN 10 ng/ml vs. fFN 50 ng/ml, they represent both a reduction in costs and QALYs. Actim Partus results in £56,030 of cost savings per QALY lost relative to fFN 50 ng/ml, which are higher than those of fFN 200 ng/ml (£25,209) and fFN 500 ng/ml (£17,025). Incremental costs and QALYs for PartoSure vs fFN 50 ng/ml are the result of an indirect comparison between Bruijn et al. and Hadzi-Lega et al., since no included study directly compares these two tests. Subject to this caveat, PartoSure would produce the same QALY loss but more cost savings than Actim Partus, relative to fFN 50 ng/ml.

Table 32 Summary of ICERs for women presenting at 30 weeks' gestation (at a level 2 hospital)*

			Versus treat all			Versus fFN 50 ng/ml		
Test	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (per QALY)	Incremental costs	Incremental QALYs	ICER (per QALY)
Actim Partus[a]	£4,891	22.010	-£1,116	-0.010	£108,319*	-£346	-0.006	£56,030*
PartoSure[b]	£4,731†	22.010†	-£1,110	-0.008	£140,587*	-£506	-0.006	£81,922*
Treat all	£6,007	22.020	£0	0	-	£770	0.004	£186,754
fFN 10 ng/ml[a]	£5,526	22.018	-£481	-0.002	£233,241*	£289	0.002	£140,267
fFN 50 ng/ml[a]	£5,237	22.016	-£770	-0.004	£186,754*	£0	0	-
fFN 200 ng/ml[a]	£4,995	22.006	-£1,012	-0.014	£73,673*	-£242	-0.010	£25,209*
fFN 500 ng/ml[a]	£4,840	21.992	-£1,167	-0.027	£42,485*	-£398	-0.023	£17,025*

Notes: ICER, incremental cost effectiveness ratio; fFN, fetal fibronectin; QALY, quality adjusted life years; [a] Bruijn et al.[45, 46]; [b] Hadzi-Lega et al.[47] for comparison with treat-all, indirect comparison between Bruijn et al. and Hadzi-Lega et al. for comparison with fFN 50 ng/ml (Bruijn et al. was used as the reference study in this case); * ICER represents the South-West quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs); † Inferred total cost and QALYs for PartoSure obtained by applying relative differences vs Actim Partus found using Hadzi-Lega et al. to Bruijn et al.

Key: ICER, incremental cost effectiveness ratio; fFN, fetal fibronectin; QALY, quality adjusted life years

Table 33 show the base case results as a full incremental analysis. The rows of the table are ordered from most to least effective testing option in terms of total QALYs. Incremental costs, QALYs, and cost-effectiveness for each test are shown in comparison to the following option in the table. For example, we see that fFN 50 ng/ml has an ICER of £56,030 relative to Actim Partus. Actim Partus, however, is itself dominated by PartoSure due to PartoSure having equal effectiveness but at a lower cost. (The ICER of fFN 50 ng/ml relative to PartoSure is £81,922, see Table 32). A graphical depiction of these results is presented in Figure 17.

Table 33 Fully incremental analysis of ICERs for women presenting at 30 weeks' gestation at a level 2 hospital

			Versus next option in the QALY ranking		
Test	Total costs	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER
Treat-all (test none)	£6,007	22.020	£481	0.002	£233,241
fFN 10 ng/ml[a]	£5,526	22.018	£289	0.002	£140,267
fFN 50 ng/ml[a]	£5,237	22.016	£346	0.006	£56,030
Actim Partus[a]	£4,891	22.010	£160	0.000	Dominated by PartoSure
PartoSure[b]	£4,731†	22.010†	-£264	0.003	-£76,873 (Dominates fFN 200 ng/ml)
fFN 200 ng/ml[a]	£4,995	22.006	£155	0.014	£11,296
fFN 500 ng/ml[a]	£4,840	21.992	-	-	-

Notes: Options have been ranked from most to least effective (in terms of QALYs). ICERs are relative to the next most effective option (i.e. the test in the row immediately below).

Key: ICER, incremental cost effectiveness ratio; fFN, fetal fibronectin; QALY, quality adjusted life years; [a] Bruijn et al.[45, 46]; [b] Hadzi-Lega et al.[47] for comparison with treat-all, indirect comparison between Bruijn et al. and Hadzi-Lega et al. for comparison with fFN 50 ng/ml (Bruijn et al. was used as the reference study in this case); † Inferred total cost and QALYs for PartoSure obtained by applying relative differences vs Actim Partus found using Hadzi-Lega et al. to Bruijn et al.

Table 34 breaks down the base case results shown in Table 32 in terms of their component discounted costs and QALYs. It should be noted, as in Table 32, that the ICER for PartoSure vs fFN 50 ng/ml is the result of an indirect comparison via Actim Partus. More specifically, the relative differences between PartoSure and Actim Partus obtained using Hadzi-Lega et al. were applied to the results for Actim Partus using Bruijn et al, and then compared to fFN 50 ng/ml.

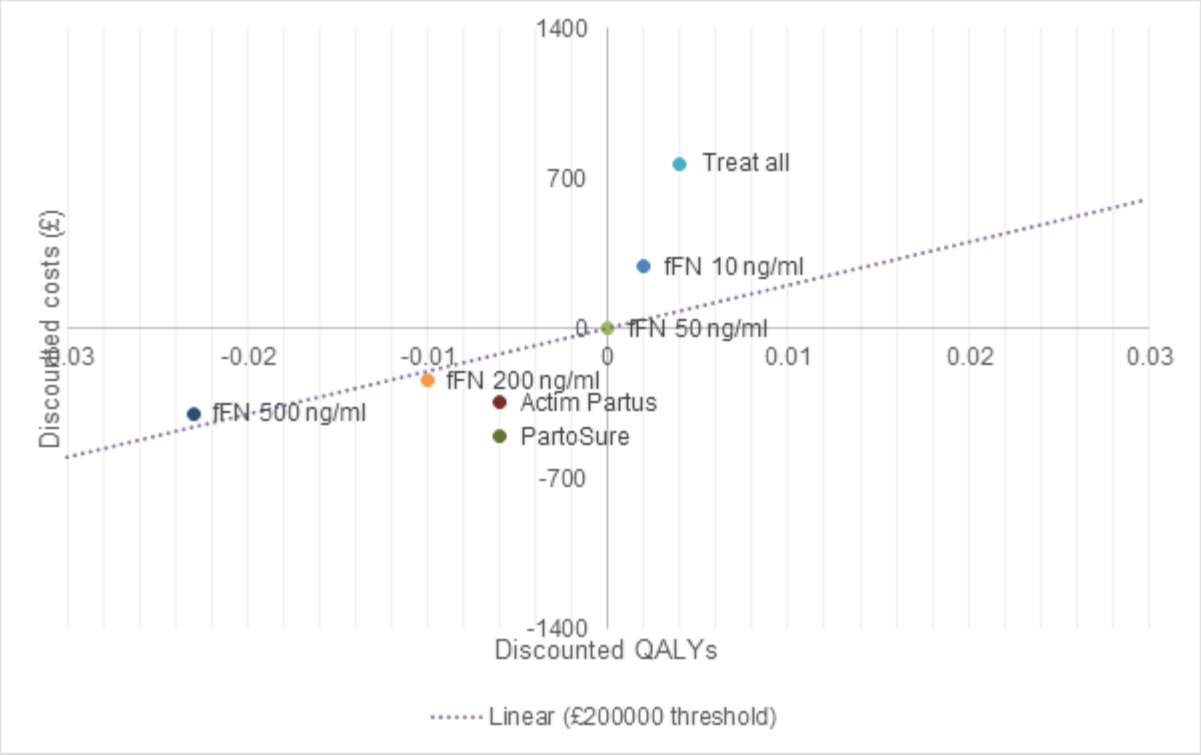
Table 34 Breakdown of base case results (discounted costs and QALYs)

			Bruijn, 2016: APOSTEL-1					Indirect comparison†
		Treat all	**fFN**	**fFN**	**fFN**	**fFN**	**Actim Partus**	**PartoSure**
	Threshold (ng/ml)		**10**	**50**	**200**	**500**		
Discounted Costs	Diagnosis	£0	£66	£66	£66	£66	£35	£52
	Treatment	£5	£3	£2	£1	£0	£1	£0
	Hospital admission	£1,325	£781	£493	£250	£95	£177	£1
	In-utero transfer	£0	£0	£0	£0	£0	£0	£0
	Neonatal IVH	£4,006	£4,008	£4,010	£4,018	£4,030	£4,015	£4,015
	Neonatal RDS	£624	£624	£625	£627	£630	£626	£626
	Neonatal death[1]	£47	£45	£43	£33	£20	£36	£36
	Total	£6,007	£5,526	£5,237	£4,995	£4,840	£4,891	£4,731

	Incremental Costs (vs. fFN 50ng/ml)	£770	£289	reference	-£242	-£398	-£346	-£506
Discounted QALYs	Surviving neonate without morbidity	22.00	22.00	22.00	22.00	22.00	22.00	22.00
	Loss new-born morbidity –IVH	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
	Loss new-born morbidity –RDS	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Loss new-born mortality	0.04	0.04	0.04	0.03	0.02	0.03	0.03
	Total	22.020	22.018	22.016	22.006	21.992	22.010	22.010
	Incremental QALYs (vs. fFN 50ng/ml)	0.004	0.002	reference	-0.010	-0.023	-0.006	-0.006
	ICER vs. fFN 50ng/ml	£186,754	£140,267	reference	£25,209*	£17,025*	£56,030*	£81,922*

Notes Key: [1] These are the neonatal hospital costs associated with those infants saved by steroid treatment; AE, adverse events; fFN, fetal fibronectin; ICER, incremental cost effectiveness ratio; QALY, quality adjusted life years; * ICER represents the South-West quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs). † Costs and QALYs are inferred values computed via an indirect comparison between the Hadzi-Lega and Bruijn studies [45-47]

Figure 17 Incremental costs and benefits of index tests against comparator (fFN @50 ng/ml)



6.2.2 Women presenting at 26 weeks' gestation (at a Level 2 hospital)

As before, ICERs should be interpreted with caution since, other than 'treat all' and fFN 10 ng/ml vs. fFN 50 ng/ml, they represent both a reduction in costs and QALYs. Incremental costs and QALYs for PartoSure vs fFN 50 ng/ml are the result of an indirect comparison between Bruijn et al. and Hadzi-Lega et al., since no included study directly compares these

two tests. As for the case of women presenting at 30 weeks, Actim Partus results in £35, 364 of cost savings per QALY lost relative to fFN 50 ng/ml, which are higher than those of fFN 200 ng/ml (£16,541) or fFN 500 ng/ml (£11,476). Based on indirect comparison, PartoSure appears to offer the same QALY loss but higher cost savings than Actim Partus, relative to fFN 50 ng/ml.

Table 35 Summary of ICERs for women presenting at 26 weeks' gestation (level 2 hospital)

			Versus treat-all			Versus fFN 50 ng/ml		
Test	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (per QALY)	Incremental costs	Incremental QALYs	ICER (per QALY)
Actim Partus[a]	£15,263	21.619	-£2,259	-0.031	£72,794*	-£658	-0.019	£35,364*
PartoSure[b]	£14,926 †	21.619 †	-£2, 266	-0.024	£95,252*	-£995	-0.019	£53,446*
Treat all	£17,522	21.650	£0	0	-	£1,600	0.012	£128,939
fFN 10 ng/ml[a]	£16,498	21.643	-£1,024	-0.006	£165,033*	£576	0.006	£92,845
fFN 50 ng/ml[a]	£15,921	21.637	-£1,600	-0.012	£128,939*	£0	0	-
fFN 200 ng/ml[a]	£15,442	21.608	-£2,080	-0.041	£50,260*	-£479	-0.029	£16,541*
fFN 500 ng/ml[a]	£15,114	21.567	-£2,408	-0.070	£29,095*	-£807	-0.070	£11,476*

Key: ICER, incremental cost effectiveness ratio; fFN, fetal fibronectin; QALY, quality adjusted life years; [a] Bruijn et al.[45, 46]; [b] Hadzi-Lega et al.[47] for comparison with treat-all, indirect comparison between Bruijn et al. and Hadzi-Lega et al. for comparison with fFN 50 ng/ml (Bruijn et al. was used as the reference study in this case); * ICER represents the South-West quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs); † Inferred total cost and QALYs for PartoSure obtained by applying relative differences vs Actim Partus found using Hadzi-Lega et al. to Bruijn et al.

Full incremental analyses are presented in Table 36, and detailed costs and QALYs are presented in Table 37. Note that diagnostic options involving wider use of treatment have become more attractive for this group of women than women presenting at older gestation ages, e.g. ICER for Treat all in Table 35, £128,939, is lower than that in Table 32, £186,754.

Table 36 Fully incremental analysis of ICERs for women presenting at 26 weeks' gestation (level 2 hospital)

			Versus next option in the QALY ranking		
Test	Total costs	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER
Treat-all (test none)	£17,522	21.650	£1,024	0.006	£165,033
fFN 10 ng/ml[a]	£16,498	21.643	£576	0.006	£92,845
fFN 50 ng/ml[a]	£15,921	21.637	£658	0.019	£35,364
Actim Partus[a]	£15,263	21.619	£337	0.000	Dominated by PartoSure
PartoSure[b]	£14,926 †	21.619 †	-£516	0.010	-£49,889 (Dominates fFN 200 ng/ml)
fFN 200 ng/ml[a]	£15,442	21.608	£328	0.041	£7,930
fFN 500 ng/ml[a]	£15,114	21.567	-	-	-

Notes: Options have been ranked from most to least effective (in terms of QALYs). ICERs are relative to the next most effective option (i.e. the test in the row immediately below).

Key: ICER, incremental cost effectiveness ratio; fFN, fetal fibronectin; QALY, quality adjusted life years; [a] Bruijn et al.[45, 46]; [b] Hadzi-Lega et al.[47] for comparison with treat-all, indirect comparison between Bruijn et al. and Hadzi-Lega et al. for comparison with fFN 50 ng/ml (Bruijn et al. was used as the reference study in this case); † Inferred total cost and QALYs for PartoSure obtained by applying relative differences vs Actim Partus found using Hadzi-Lega et al. to Bruijn et al.

Table 37 Breakdown of discounted costs and QALYs for women presenting at 26 weeks' gestation

		Bruijn, 2016: APOSTEL-1						Indirect comparison†
		Treat all	**fFN**	**fFN**	**fFN**	**fFN**	**Actim Partus**	**PartoSure**
	Threshold (ng/ml)		**10**	**50**	**200**	**500**		
Discounted Costs	Diagnosis	£0	£66	£66	£66	£66	£35	£52
	Treatment	£367	£216	£136	£69	£0	£49	£0
	Hospital admission	£1,325	£781	£493	£250	£95	£177	£1
	In-utero transfer	£965	£569	£359	£182	£69	£129	£1
	Neonatal IVH	£5,232	£5,235	£5,237	£5,248	£5,264	£5,244	£5,244
	Neonatal RDS	£9,467	£9,473	£9,480	£9,509	£9,552	£9,499	£9,499
	Neonatal death[1]	£166	£158	£151	£118	£70	£130	£130
	Total	£17,522	£16,498	£15,921	£15,442	£15,114	£15,263	£14,926
	Incremental Costs (vs. fFN 50ng/ml)	£1,600	£576	reference	-£479	-£807	-£658	-£995
Discounted QALYs	Surviving neonate without morbidity	21.55	21.55	21.55	21.55	21.55	21.55	21.55
	Loss new-born morbidity –IVH	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03
	Loss new-born morbidity –RDS	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01

	Loss new-born mortality	0.13	0.13	0.12	0.10	0.06	0.10	0.10
	Total	21.650	21.643	21.637	21.608	21.567	21.619	21.619
	Incremental QALYs (vs. fFN 50ng/ml)	0.012	0.006	reference	-0.029	-0.070	-0.019	-0.019
	ICER vs. fFN 50ng/ml	£128,939	£92,845	**reference**	£16,541*	£11,476*	£35,364*	£53,446*

Notes [1] These are the neonatal hospital costs associated with those infants saved by steroid treatment; * ICER represents the South-West quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs). † Costs and QALYs are inferred values computed via an indirect comparison between the Hadzi-Lega and Bruijn studies [45-47]

Key: AE, adverse events; fFN, fetal fibronectin; ICER, incremental cost effectiveness ratio; QALY, quality adjusted life years

6.2.3 Women presenting at 33 weeks' gestation (at a Level 2 hospital)

Qualitatively similar results to those described before for women presenting at 26 and 30 weeks were found for women presenting at 33 weeks' gestation. At £97,069, Actim Partus saves more costs per QALY lost relative to fFN 50 ng/ml than fFN 200 ng/ml (£43,781) and fFN 500 ng/ml (£29,631), while Treat all and fFN 10 ng/ml both have incremental costs per QALY gained that are above £200,000. Based on indirect comparison, PartoSure appears to dominate Actim Partus as it results in the same amount of QALYs and lower costs. Table 39 presents the summary results for each test relative to the comparators, Table 40 presents the fully incremental analyses and Table 41 the detailed costs and QALY elements.

Table 38 Summary of ICERs for women presenting at 33 weeks' gestation (level 2 hospital)

			Versus treat-all			Versus fFN 50 ng/ml		
Test	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (per QALY)	Incremental costs	Incremental QALYs	ICER (per QALY)
Actim Partus[a]	£2,716	22.096	-£1,117	-0.006	£187,479*	-£347	-0.004	£97,069*
PartoSure[b]	£2,556 †	22.096 †	-£1,111	-0.005	£243,269*	-£507	-0.004	£141,838*
Treat all	£3,833	22.102	£0	0	-	£770	0.002	£323,093
fFN 10 ng/ml[a]	£3,352	22.101	-£481	-0.001	£403,469*	£289	0.001	£242,716
fFN 50 ng/ml[a]	£3,063	22.100	-£770	-0.002	£323,093*	£0	0	-
fFN 200 ng/ml[a]	£2,820	22.094	-£1,013	-0.008	£127,575*	-£243	-0.006	£43,781*
fFN 500 ng/ml[a]	£2,663	22.086	-£1,170	-0.016	£73,650*	-£400	-0.014	£29,631*

Key: ICER, incremental cost effectiveness ratio; fFN, fetal fibronectin; QALY, quality adjusted life years; [a] Bruijn et al.[45, 46]; [b] Hadzi-Lega et al.[47] for comparison with treat-all, indirect comparison between Bruijn et al. and Hadzi-Lega et al. for comparison with fFN 50 ng/ml (Bruijn et al. was used as the reference study in this case); * ICER represents the South-West quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs); † Inferred total cost and QALYs for PartoSure obtained by applying relative differences vs Actim Partus found using Hadzi-Lega et al. to Bruijn et al.

Full incremental analyses are presented in Table 39, and detailed costs and QALYs are presented in Table 40. Diagnostic options involving wider use of treatment have become less attractive for this group of women than women presenting at younger gestation ages, e.g. ICER for Treat all in Table 39, £323,093 is higher than those in Table 35, £128,939, and Table 32, £186,754.

Table 39 Fully incremental analysis of ICERs for women presenting at 33 weeks' gestation (level 2 hospital)

			Versus next option in the QALY ranking		
Test	Total costs	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER
Treat-all (test none)	£3,833	22.102	£481	0.001	£403,469
fFN 10 ng/ml[a]	£3,352	22.101	£289	0.001	£242,716
fFN 50 ng/ml[a]	£3,063	22.100	£347	0.004	£97,069
Actim Partus[a]	£2,716	22.096	£160	0.000	Dominated by PartoSure
PartoSure[b]	£2,556 †	22.096 †	-£264	0.002	-£132,721 (Dominates fFN 200 ng/ml)
fFN 200 ng/ml[a]	£2,820	22.094	£157	0.008	£19,725
fFN 500 ng/ml[a]	£2,663	22.086	-	-	-

Notes: Options have been ranked from most to least effective (in terms of QALYs). ICERs are relative to the next most effective option (i.e. the test in the row immediately below).

Key: ICER, incremental cost effectiveness ratio; fFN, fetal fibronectin; QALY, quality adjusted life years; [a] Bruijn et al.[45, 46]; [b] Hadzi-Lega et al.[47] for comparison with treat-all, indirect comparison between Bruijn et al. and Hadzi-Lega et al. for comparison with fFN 50 ng/ml (Bruijn et al. was used as the reference study in this case); † Inferred total cost and QALYs for PartoSure obtained by applying relative differences vs Actim Partus found using Hadzi-Lega et al. to Bruijn et al.

Table 40 Breakdown of discounted costs and QALYs for women presenting at 33 weeks' gestation

			Bruijn, 2016: APOSTEL-1					Indirect comparison†
		Treat all	fFN	fFN	fFN	fFN	Actim Partus	PartoSure
	Threshold (ng/ml)		10	50	200	500		
Discounted Costs	Diagnosis	£0	£66	£66	£66	£66	£35	£52
	Treatment	£5	£3	£2	£1	£0	£1	£0
	Hospital admission	£1,325	£781	£493	£250	£95	£177	£1
	In-utero transfer	£0	£0	£0	£0	£0	£0	£0
	Neonatal IVH[2]	£2,477	£2,478	£2,479	£2,484	£2,492	£2,482	£2,482
	Neonatal RDS[2]	£0	£0	£0	£0	£0	£0	£0
	Neonatal death[1]	£26	£25	£24	£19	£11	£21	£21
	Total	£3,833	£3,352	£3,063	£2,820	£2,663	£2,716	£2,556
	Incremental Costs (vs. fFN 50ng/ml)	£770	£289	reference	-£243	-£400	-£347	-£507
Discounted QALYs	Surviving neonate without morbidity	22.09	22.09	22.09	22.09	22.09	22.09	22.09
	Loss new-born morbidity –IVH	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	Loss new-born morbidity –RDS	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Loss new-born mortality	0.03	0.02	0.02	0.02	0.01	0.02	0.02
	Total	22.102	22.101	22.100	22.094	22.086	22.096	22.096
	Incremental QALYs (vs. fFN 50ng/ml)	0.002	0.001	reference	-0.006	-0.014	-0.004	-0.004
	ICER vs. fFN 50ng/ml	£323,093	£242,716	-	£43,781*	£29,631*	£97,069*	£141,838*

Notes Key: [1] These are the neonatal hospital costs associated with those infants saved by steroid treatment; AE, adverse events; fFN, fetal fibronectin; ICER, incremental cost effectiveness ratio; QALY, quality adjusted life years; * ICER represents the South-West quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs). † Costs and QALYs are inferred values computed via an indirect comparison between the Hadzi-Lega and Bruijn studies [45-47]

6.2.4 Tornado Analysis

For the tornado analysis, the parameter base-case values were increased and decreased by 20% (the upper and the lower variations, respectively) and the ICERs vs the comparator of fFN 50 ng/ml were plotted, with the intersection of the vertical and the horizontal axes at the ICER base-case. The tornado plots for each of the interventions in the Bruijn 2016: APOSTEL-1 study,[46] the Hadzi-Lega 2017 study[47] are presented below.

There is a consistent pattern across all comparisons, as depicted in Figure 18, Figure 19, Figure 20, Figure 21, Figure 22, and Figure 23. The results are sensitive to the health related quality of life (state utility) of preterm survivors. Much less influential are the cost of hospital admission, the prevalence of preterm birth within 7 days, the effectiveness of steroid

treatment and the baseline mortality risks. Other parameter values appear to have no discernible influence on the results.

Figure 18 Tornado diagram of most influential model parameter -Treat all

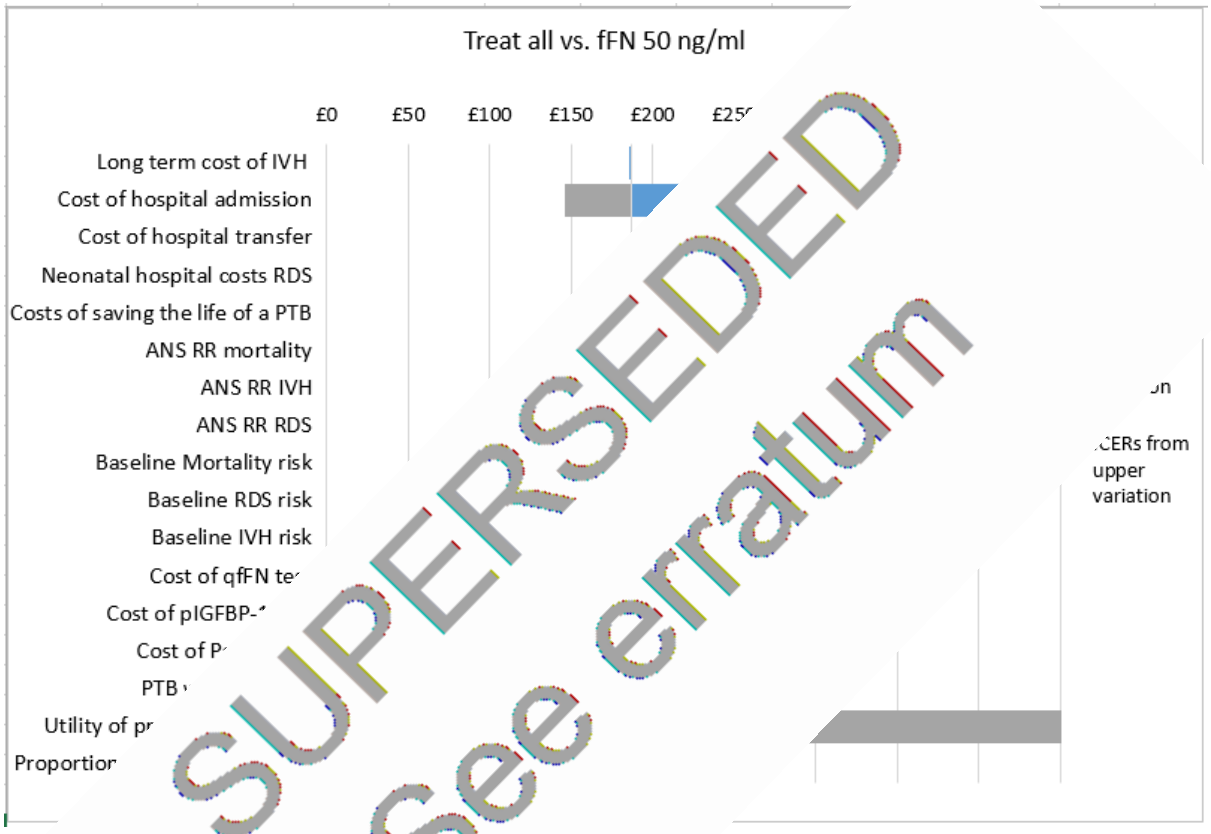


Figure 19 Tornado diagram of most influential model parameter values –fFN 10ng/ml

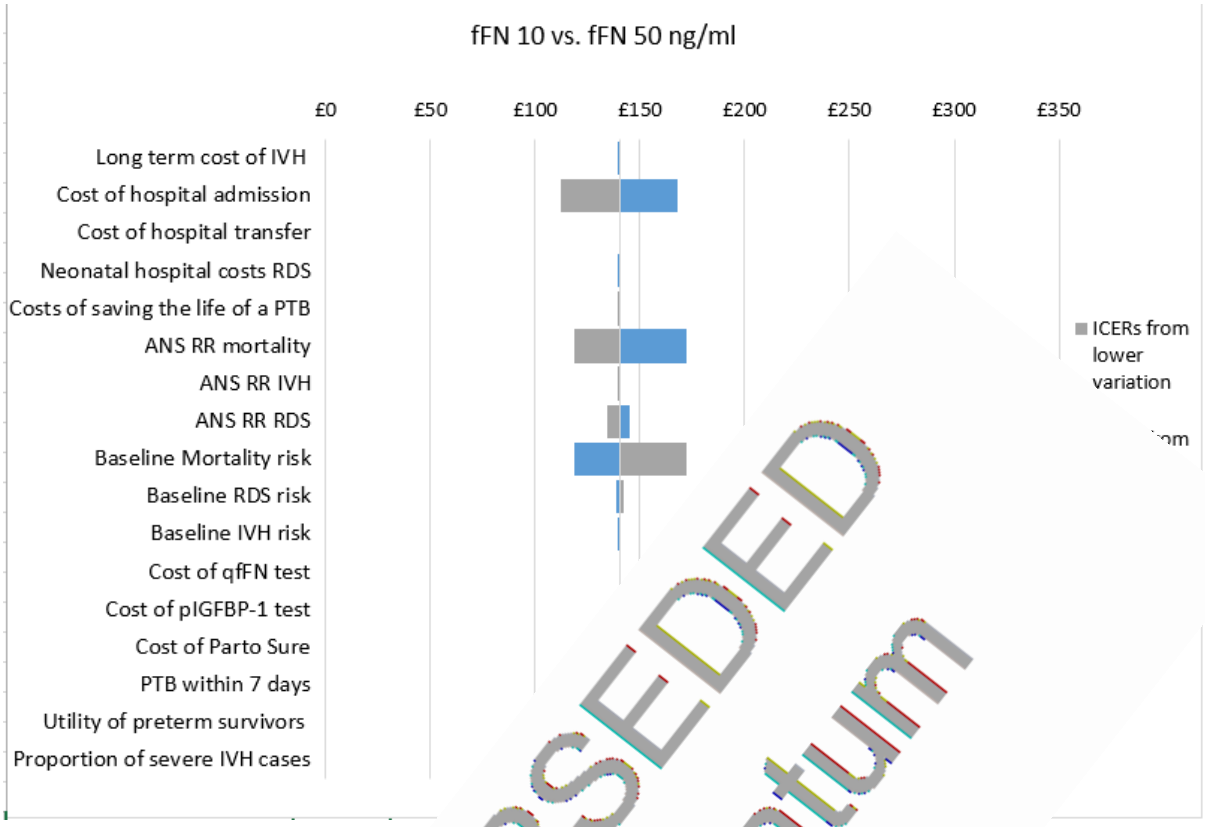


Figure 20 Tornado diagram –fFN 200ng/ml

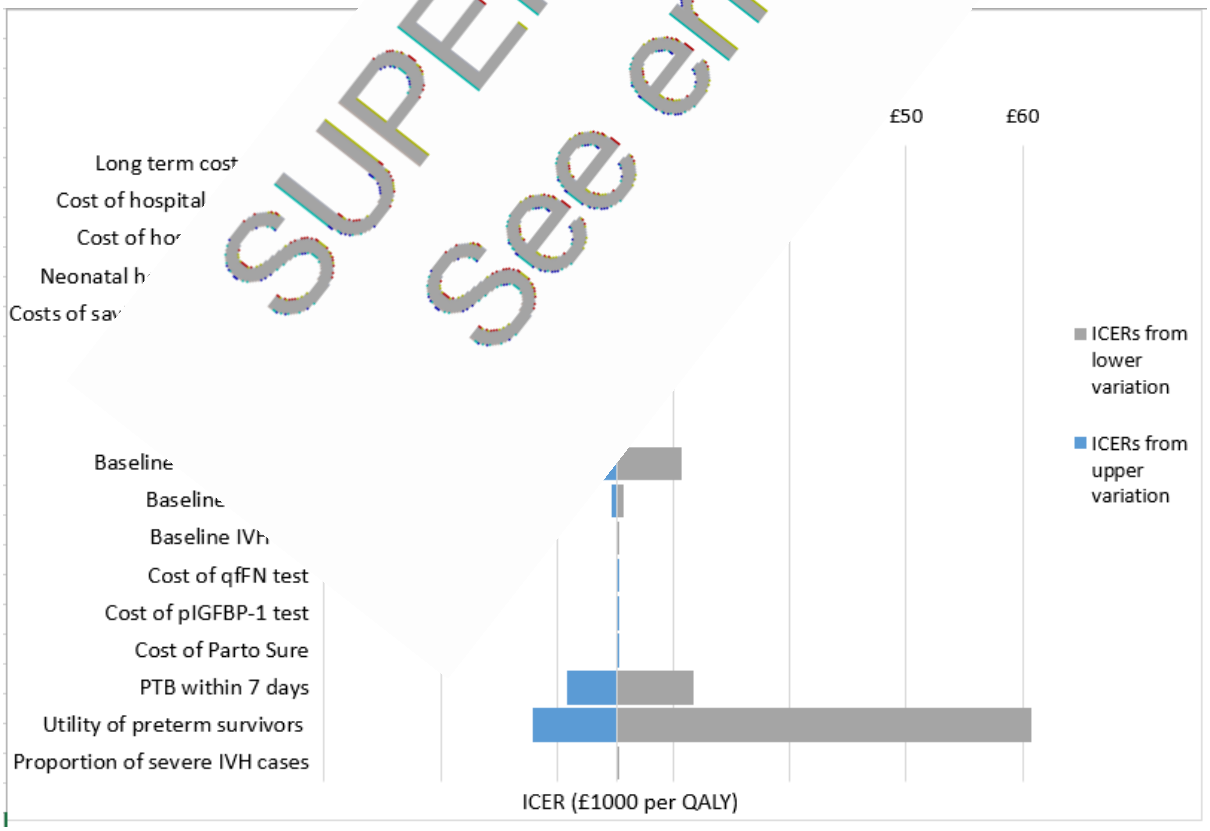


Figure 21 Tornado diagram of most influential model parameter values –fFN 500ng/ml

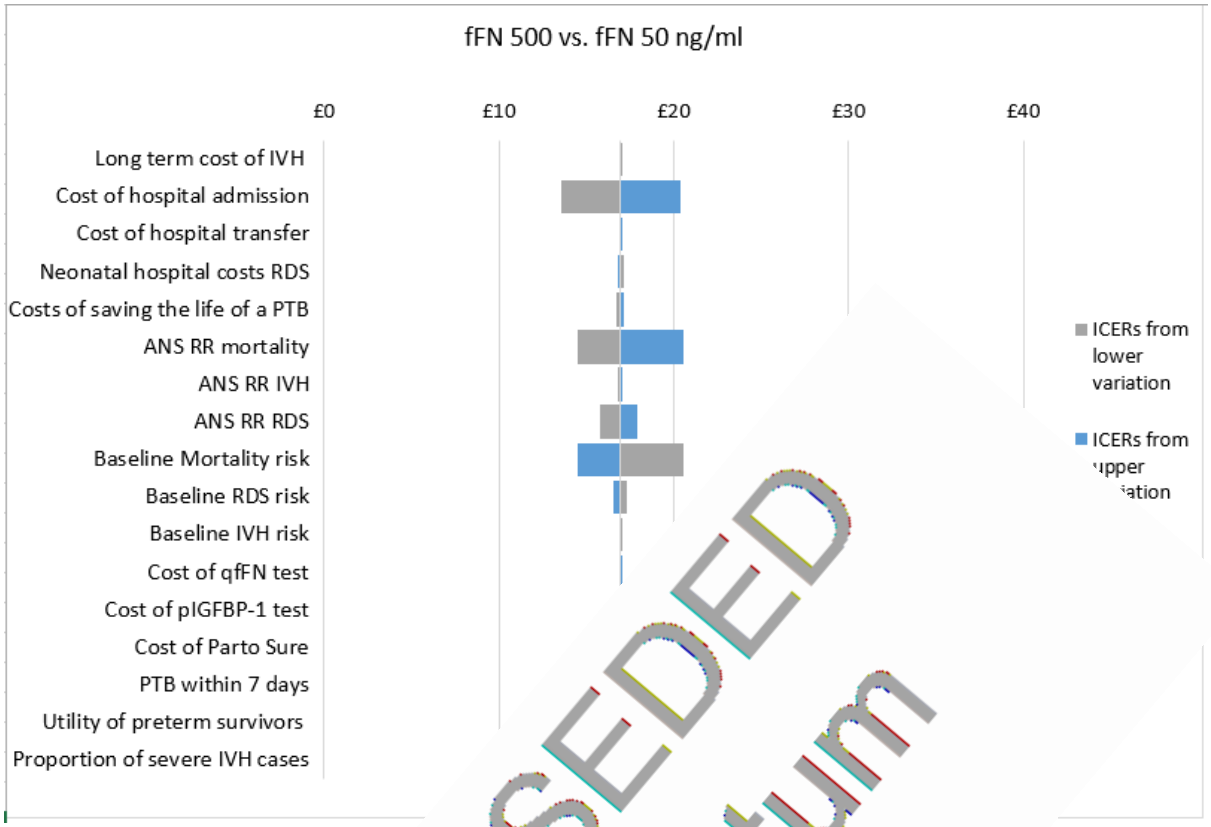
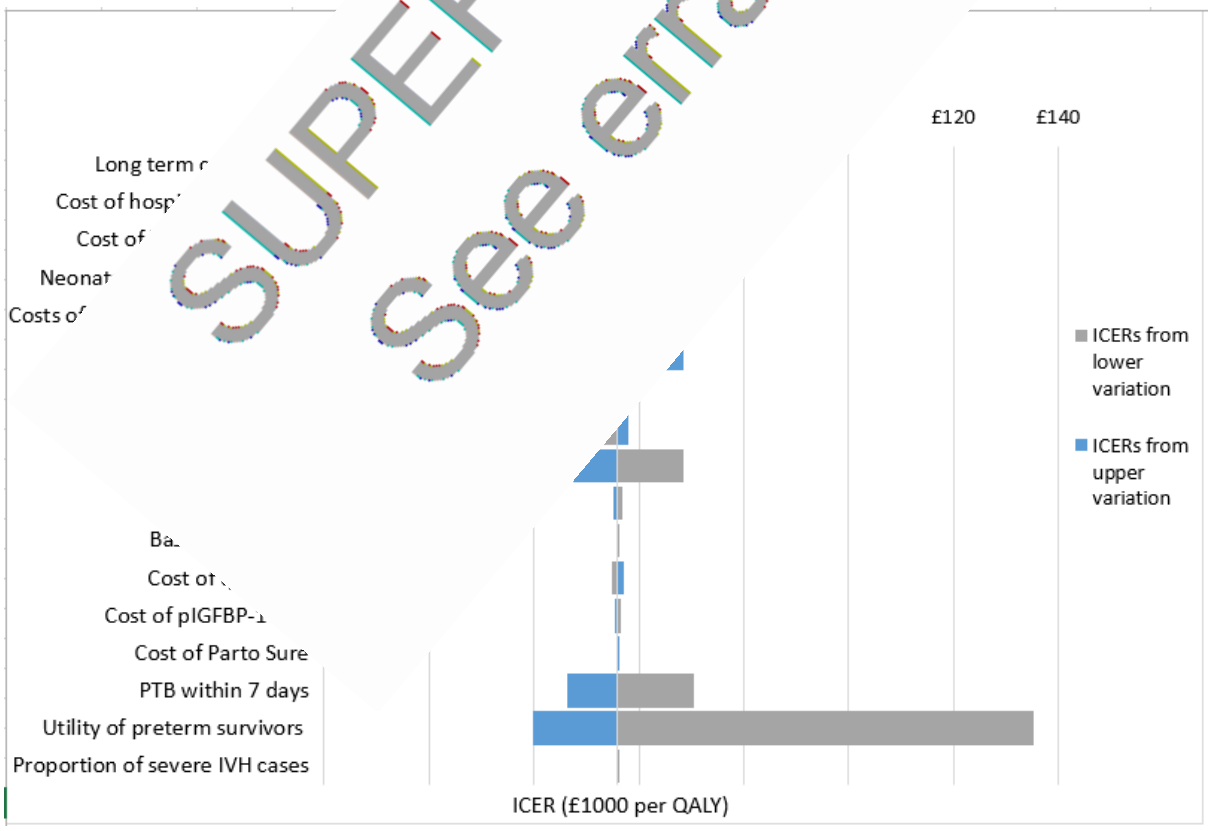
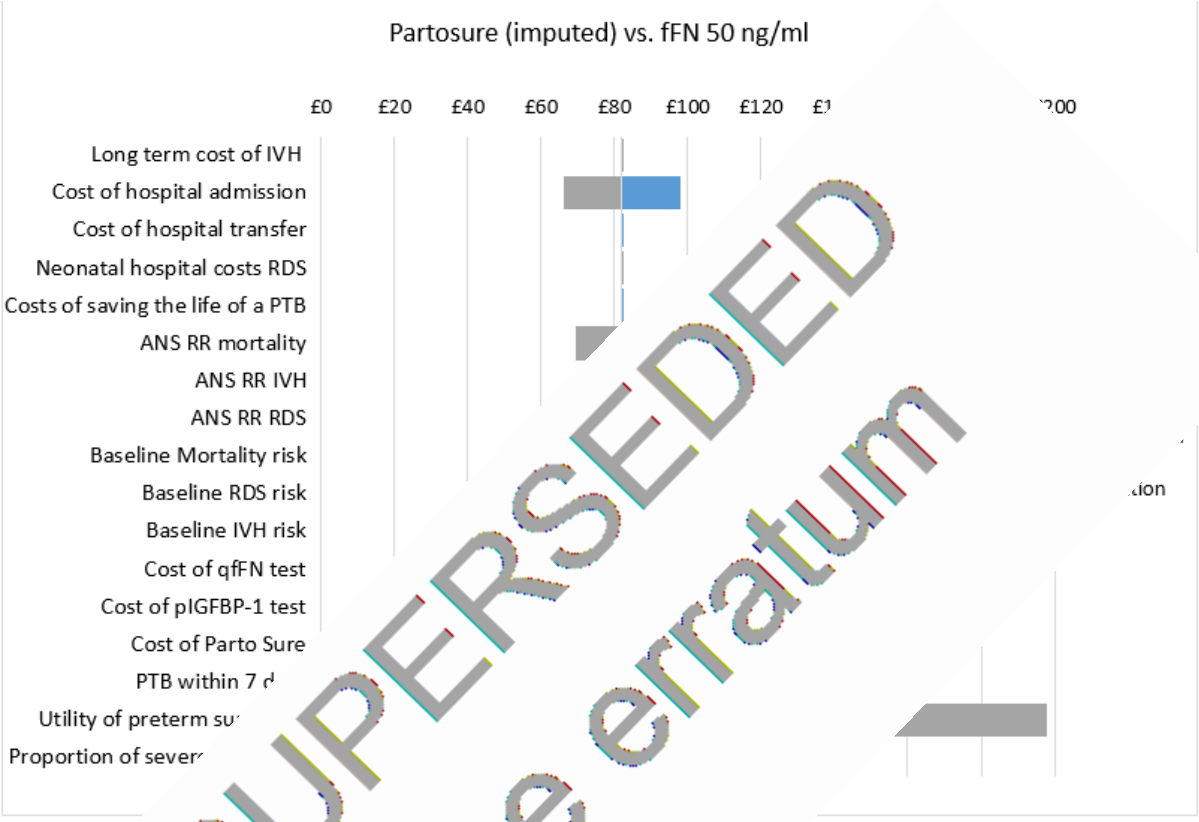


Figure 22 Tornado diagra –Actim Partus



Hadzi-Lega, 2017

Figure 23 Tornado diagram of most influential model parameter values –PartoSure



6.2.5

At ...um Partus has a probability of being c... ...esenting at gestational ages 33, 30 and 26 week... , and Figure 26). PartoSure has probabilities of being c... ...%, respectively, but these values are based on indirect comp... ...trength of evidence than for other diagnostic options.

Treat all has a prob... ...g cost-effective of 0% at a £20,000 willingness to pay per QALY threshold for all ..., and only becomes the option with the highest likelihood of being cost-effective for women presenting at 26 weeks' gestation at a willingness to pay per QALY threshold above £180,000 (Figure 26).

Figure 24 Probabilistic analysis –women presenting at 33 weeks' gestation

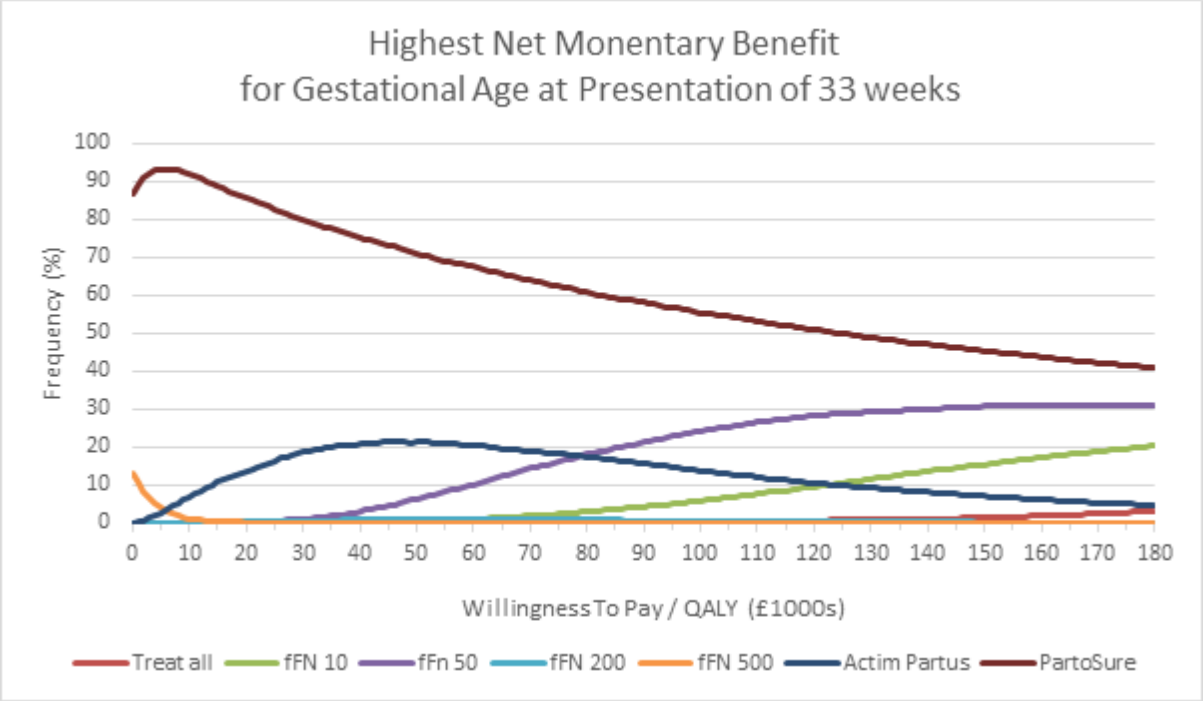


Figure 25 Probabilistic analysis –women presenting at 30 weeks' gestation

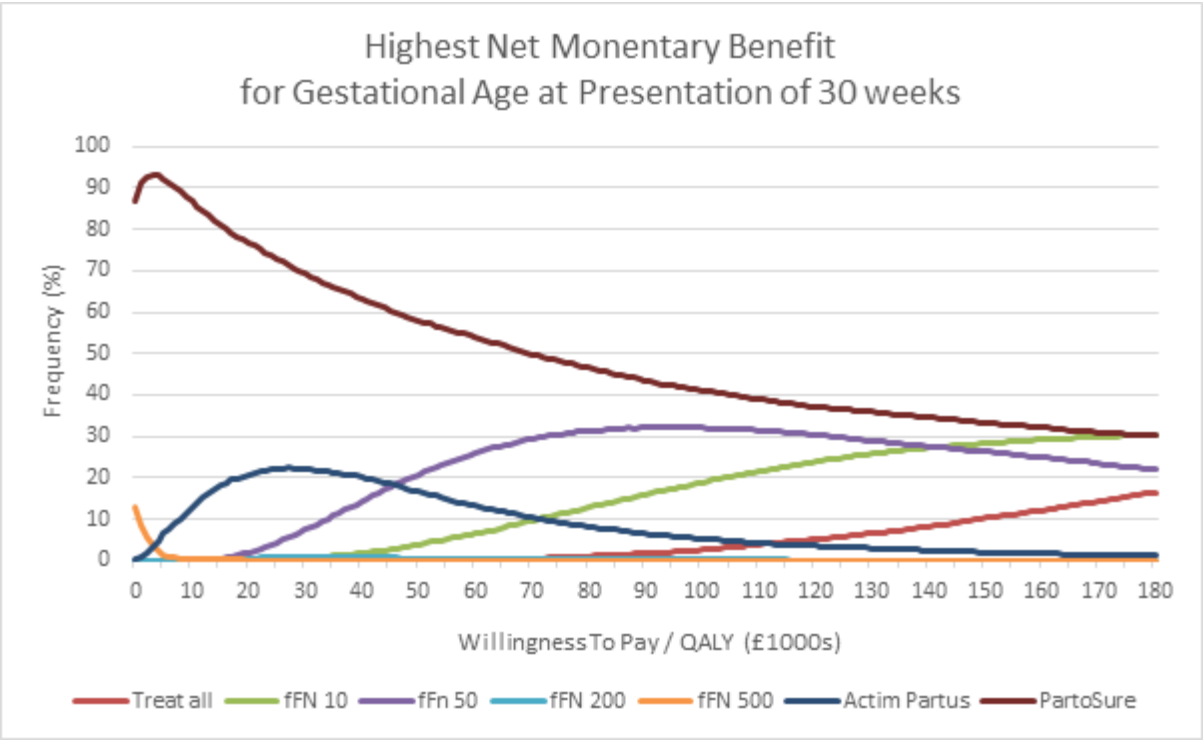
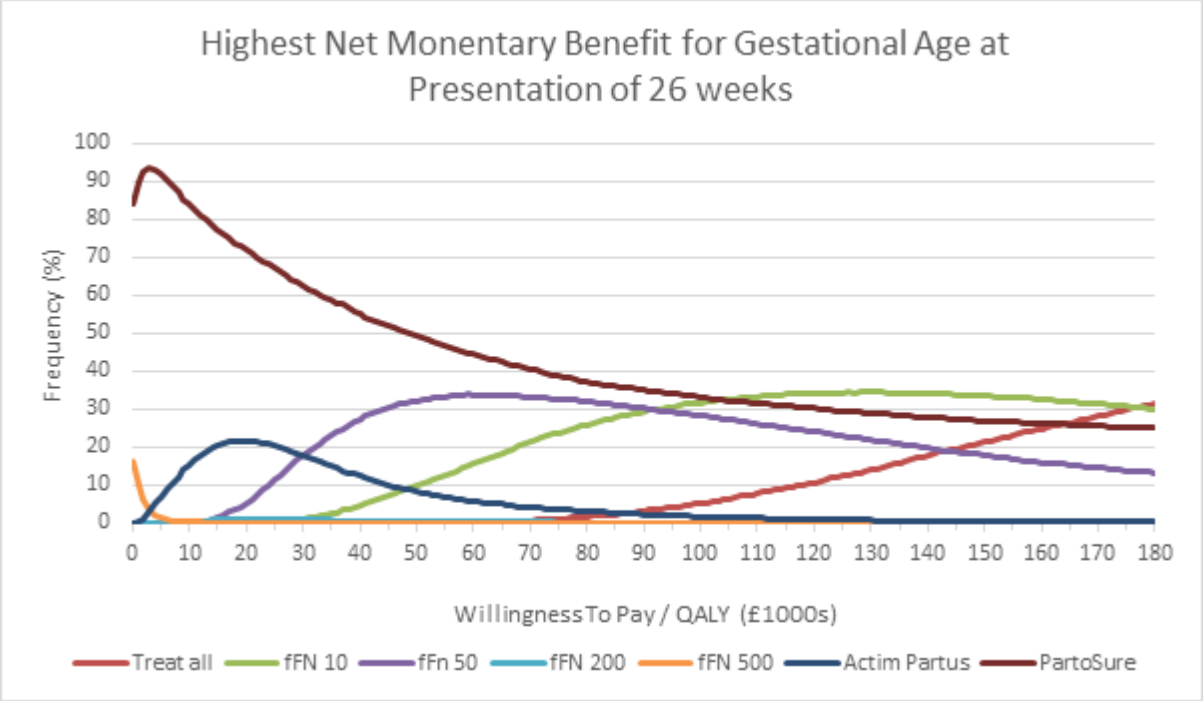


Figure 26 Probabilistic analysis –women presenting at 26 weeks' gestation



6.2.6 Scenario Analyses

6.2.6.1 Alternative diagnostic accuracy data

Using the diagnostic test accuracy results of the study by Cooper and colleagues,[1] which did not include PartoSure, suggests that the fFN 50 ng/ml test provides lower costs with equal health benefit when compared to Actim Partus. The option of treating all women compared with fFN 50 ng/ml yields an ICER of £34,508 per QALY (Table 41).

Table 41 Results for Actim Partus and no-testing vs fFN 50 ng/ml using data from Cooper 2012;[1] presenting at 30 weeks gestation (level 2 hospital)

		Cooper 2012[1]		
		Treat all	Actim Partus	fFN 50 ng/ml
Discounted Costs	Diagnosis	£0	£35	£66
	Medication	£5	£1	£1
	Admission	£1,325	£373	£171
	Transfer	£0	£0	£0
	RDS	£4,006	£4,034	£4,034
	IVH	£624	£630	£630
	Neonatal death	£47	£16	£16
	Total	**£6,007**	**£5,090**	**£4,917**
	Incremental costs vs. fFN 50	£1,090	£173	
Discounted QALYs	Baseline w/o morbidity	22.00	22.00	22.00
	RDS	-0.02	-0.02	-0.02
	IVH	0.00	0.00	0.00
	Newborn mortality	0.04	0.01	0.01
	Total	**22.02**	**21.99**	**21.99**
	Incremental QALYs vs. fFN 50	0.03	0.00	
ICER vs fFN 50 ng/ml		**£34,508**	**Dominated**	**-**

Key: fFN, fetal fibronectin; ICER, incremental cost effectiveness ratio; QALY, quality adjusted life year

[redacted]

Table 42 Results for fFN (various thresholds) vs fFN 50 ng/ml using data from Abbott 2013; presenting at 30 weeks gestation (level 2 hospital)

		Abbott 2013[102]				
		Treat all	fFN 10	fFN 200	fFN 500	fFN 50
Discounted Costs	Diagnosis					
	Medication					
	Admission					
	Transfer					
	RDS					
	IVH					
	Neonatal death					
	Total					
	Incremental costs vs. fFN 50					
Discounted QALYs	Baseline w/o morbidity					
	RDS					
	IVH					
	Newborn mortality					
	Total					
	Incremental QALYs vs. fFN 50					
ICER vs fFN 50 ng/ml						

Key: fFN, fetal fibronectin; ICER, incremental cost effectiveness ratio; QALY, quality adjusted life year
* ICER represents the South-West quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

When the diagnostic accuracy data from the meta-analysis of the four studies that compared Actim Partus with fFN 50ng/ml were used,[45, 46, 53, 58, 59] Actim Partus was dominant over fFN 50ng/ml as it resulted in cost savings of £41 per woman and health benefits of 0.01 more QALYs per woman. Treat all increased costs and QALYs, and resulted in an ICER of £70,468.

Table 43 Results for no-testing and Actim Partus vs fFN 50 ng/ml using data from meta-analysis; presenting at 30 weeks gestation (level 2 hospital)

		Meta-analysis[45, 46, 53, 58, 59]		
		Treat all	Actim Partus	fFN 50 ng/ml
Discounted Costs	Diagnosis	£0	£35	£66
	Medication	£5	£1	£1
	Admission	£1,325	£195	£204
	Transfer	£0	£0	£0
	RDS	£4,006	£4,013	£4,019
	IVH	£624	£626	£627
	Neonatal death	£47	£39	£32
	Total	**£6,007**	**£4,908**	**£4,949**
	Incremental costs vs fFN 50 ng/ml	£1,058	-£41	
Discounted QALYs	Baseline w/o morbidity	22.00	22.00	22.00
	RDS	-0.02	-0.02	-0.02
	IVH	0.00	0.00	0.00
	Newborn mortality	0.04	0.04	0.03
	Mother	0.00	0.00	0.00
	Total	**22.02**	**22.01**	**22.00**
	Incremental costs vs fFN 50 ng/ml	0.02	0.01	
ICER (relative to fFN 50 ng/ml)		**£70,468**	**Dominant**	**-**

Key: fFN, fetal fibronectin; ICER, incremental cost effectiveness ratio; QALY, quality adjusted life year

6.2.6.2 Other scenarios

Including the negative impact on QALY outcomes the effect of an infant's death on mothers that is assumed to last for 10 years, favours options that involve more use of ANS treatment. That is, in Table 44, Table 45, and Table 46 the ICER for Treat all and fFN 10ng/ml under the column 'with maternal QALYs for 10 years' is lower than in the Base case column. When we limit the analytical horizon to the time of deliver, the assessment becomes in effect a cost-minimisation analysis since our model does not account for health related quality of life outcomes of mother during the antenatal period. In this scenario among women presenting at age 30 weeks, PartoSure is the least costly option with a £507 reduction in costs per woman, followed by fFN 500 ng/ml, £400, and Actim Partus, £347. As discussed before the values for PartoSure need to be considered with caution.

When we allow for partial benefits of ANS given earlier than 7 days before birth, the ICER for fFN 10 ng/ml and Treat all are £24,420 and £41,625, respectively, among women presenting at 30 weeks; as for the rest, only PartoSure results in savings per QALY lost >£20,000 relative to fFN 50 ng/ml (Table 44). Similar results apply to women presenting at gestational

ages of 26 and 33 weeks, except for the result that Actim Partus saves £24,532 in healthcare costs per QALY lost among women aged 33 weeks.

Of note, among women aged 26 weeks, excluding the neonatal hospitalisation costs associated with saving an infant's life by timely administration of ANS has the effect of halving the ICERs relative to fFN 50ng/ml. Therefore, this favours treatment-intensive options Treat All and fFN 10ng/ml, which now have an ICER of £61,791 and £46,358, respectively; other options, are favoured by the change, but all now save less than £20,000 per QALY lost relative to fFN 50 ng/ml, except for PartoSure, which saves £26,988 per lost QALY (Table 45).

Table 44 Incremental cost –effectiveness ratios (ICERs) vs fFN 50 ng/ml for women presenting at 30 weeks (level 2 hospital)

Option	Base case	With maternal QALYs for 10 years	Limiting the analysis to delivery (additional cost only)	Limiting the analysis to first year after birth	ANS earlier than 7 days before preterm delivery has partial benefits	Excluding additional neonatal hospital costs of death	Women presenting at level 3 hospital	Applying costs and disutilities of AEs to all AEs
Treat all	£186,754	£111,813	£770	£4,930,356	£41,625	£185,771	£186,754	£175,158
fFN 10 ng/ml[a]	£140,267	£74,564	£289	£3,704,141	£24,420	£139,284	£140,267	£131,558
fFN 200 ng/ml[a]	£25,209*	£18,968*	-£243	£669,219*	£9,728*	£24,226*	£25,209*	£23,664*
fFN 500 ng/ml[a]	£17,025*	£13,347*	-£400	£453,340*	£7,428*	£16,042*	£17,025*	£15,968*
Actim Partus[a]	£56,030*	£38,200*	-£347	£1,482,175*	£16,662*	£55,046*	£56,030*	£52,551*
PartoSure[b]	£81,922*	£81,893*	-£507	£2,165,156*	£128,506*	£80,939*	£81,922*	£76,836*

Key: ANS, antenatal corticosteroids; AE, adverse event; ICER, incremental cost effectiveness ratio; fFN, fetal fibronectin; QALY, quality adjusted life years; [a] Bruijn et al.[45, 46]; [b] Indirect comparison between Bruijn et al. and Hadzi-Lega et al.[47] (Bruijn et al. was used as the reference study in this case); * ICER represents the South-West quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

Table 45 Incremental cost –effectiveness ratios (ICERs) vs fFN 50 ng/ml for women presenting at 26 weeks (level 2 hospital)

Option	Base case	With maternal QALYs for 10 years	Limiting the analysis to delivery (additional cost only)	Limiting the analysis to first year after birth	ANS earlier than 7 days before preterm delivery has partial benefits	Excluding additional neonatal hospital costs of death	Women presenting at level 3 hospital	Applying costs and disutilities of AEs to all AEs
Treat all	£128,939	£72,006	£1,603	£3,422,534	£41,153	£127,779	£61,791	£117,174
fFN 10 ng/ml[a]	£92,845	£45,524	£578	£2,470,464	£23,957	£91,685	£46,358	£84,373
fFN 200 ng/ml[a]	£16,541*	£11,916*	-£486	£457,751*	£8,557*	£15,381*	£8,161*	£15,032*
fFN 500 ng/ml[a]	£11,476*	£8,660*	-£824	£324,143*	£6,576*	£10,316*	£5,444*	£10,429*
Actim Partus[a]	£35,364*	£22,807*	-£663	£954,254*	£14,629*	£34,204*	£18,392*	£32,137*
PartoSure[b]	£53,446*	£53,424*	-£1000	£1,431,224*	£68,857*	£52,287*	£26,988*	£48,570*

Key: ANS, antenatal corticosteroids; AE, adverse event; ICER, incremental cost effectiveness ratio; fFN, fetal fibronectin; QALY, quality adjusted life years; [a] Bruijn et al.[45, 46]; [b] Indirect comparison between Bruijn et al. and Hadzi-Lega et al.[47] (Bruijn et al. was used as the reference study in this case); * ICER represents the South-West quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

Table 46 Incremental cost –effectiveness ratios (ICERs) vs fFN 50 ng/ml for women presenting at 33 weeks (level 2 hospital)

Option	Base case	With maternal QALYs for 10 years	Limiting the analysis to delivery (additional cost only)	Limiting the analysis to first year after birth	ANS earlier than 7 days before preterm delivery has partial benefits	Excluding additional neonatal hospital costs of death	Women presenting at level 3 hospital	Applying costs and disutilities of AEs to all AEs
Treat all	£323,093	£194,770	£770	£8,522,367	£59,091	£322,126	£323,093	£306,507
fFN 10 ng/ml[a]	£242,716	£130,060	£289	£6,402,235	£34,621	£241,750	£242,716	£230,256
fFN 200 ng/ml[a]	£43,781*	£33,081*	-£243	£1,154,838*	£14,902*	£42,815*	£43,781*	£41,534*
fFN 500 ng/ml[a]	£29,631*	£23,314*	-£400	£781,581*	£11,654*	£28,664*	£29,631*	£28,110*
Actim Partus[a]	£97,069*	£66,541*	-£347	£2,560,443*	£24,532*	£96,103*	£97,069*	£92,086*
PartoSure[b]	£141,838*	£141,788*	-£507	£3,741,321*	£267,481*	£140,871*	£141,838*	£134,556*

Key: ANS, antenatal corticosteroids; AE, adverse event; ICER, incremental cost effectiveness ratio; fFN, fetal fibronectin; QALY, quality adjusted life years; [a] Bruijn et al.[45, 46]; [b] Indirect comparison between Bruijn et al. and Hadzi-Lega et al.[47] (Bruijn et al. was used as the reference study in this case); * ICER represents the South-West quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

Table 47 Summary of ICERs for a woman presenting at 30 weeks' gestation (level 2 hospital), including QALY losses to the mother for 10 years in case of infant mortality

			Versus treat-all			Versus fFN 50 ng/ml		
Test	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (per QALY)	Incremental costs	Incremental QALYs	ICER (per QALY)
Actim Partus[a]	£4,891	22.016	-£1,116	-0.016	£69,968*	-£346	-0.009	£38,200*
PartoSure[b]	£4,731	22.019	-£1,110	-0.013	£88,385*	-£506	-0.006	£81,893*
fFN 10 ng/ml[a]	£5,526	22.029	-£481	-0.003	£159,831*	£289	0.004	£74,564
fFN 50 ng/ml[a]	£5,237	22.025	-£770	-0.007	£111,813*	£0	0.000	-
fFN 200 ng/ml[a]	£4,995	22.012	-£1,012	-0.020	£51,469*	-£242	-0.013	£18,968*
fFN 500 ng/ml[a]	£4,840	21.995	-£1,167	-0.037	£31,829*	-£398	-0.030	£13,347*

Key: ICER, incremental cost effectiveness ratio; fFN, fetal fibronectin; QALY, quality adjusted life years; [a] Bruijn et al.[45, 46]; [b] Hadzi-Lega et al.[47] for comparison with treat-all, indirect comparison between Bruijn et al. and Hadzi-Lega et al. for comparison with fFN 50 ng/ml (Bruijn et al. was used as the reference study in this case); * ICER represents the South-West quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs); † Inferred total cost and QALYs for PartoSure obtained by applying relative differences vs Actim Partus found using Hadzi-Lega et al. to Bruijn et al.

Table 48 Summary of ICERs for a woman presenting at 30 weeks' gestation (level 2 hospital), including lifetime QALY losses to the mother in case of infant mortality

			Versus treat-all			Versus fFN 50 ng/ml		
Test	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (per QALY)	Incremental costs	Incremental QALYs	ICER (per QALY)
Actim Partus[a]	£4,891	22.027	-£1,116	-0.025	£43,954*	-£346	-0.014	£24,936*
PartoSure[b]	£4,731 †	22.035 †	-£1,110	-0.020	£54,536*	-£506	-0.006	£81,844*
fFN 10 ng/ml[a]	£5,526	22.048	-£481	-0.005	£104,731*	£289	0.007	£41,820
fFN 50 ng/ml[a]	£5,237	22.041	-£770	-0.011	£66,923*	£0	0.000	-
fFN 200 ng/ml[a]	£4,995	22.023	-£1,012	-0.030	£34,226*	-£242	-0.018	£13,416*
fFN 500 ng/ml[a]	£4,840	22.000	-£1,167	-0.052	£22,426*	-£398	-0.041	£9,806*

Key: ICER, incremental cost effectiveness ratio; fFN, fetal fibronectin; QALY, quality adjusted life years; [a] Bruijn et al.[45, 46]; [b] Hadzi-Lega et al.[47] for comparison with treat-all, indirect comparison between Bruijn et al. and Hadzi-Lega et al. for comparison with fFN 50 ng/ml (Bruijn et al. was used as the reference study in this case); * ICER represents the South-West quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs); † Inferred total cost and QALYs for PartoSure obtained by applying relative differences vs Actim Partus found using Hadzi-Lega et al. to Bruijn et al.

7 Assessment of factors relevant to the NHS and other parties

- Practical considerations when performing these tests highlighted to us by our advising obstetricians were as follows:
    - The quantitative fFN swab can be collected at an appropriate time (e.g. when a woman is being examined) and if the clinician decides it is best to 'wait and see' for a few hours, the sample can be stored. Using the pack to collect the sample is free, it is only when the cassette is opened to run the test that a cost is incurred by the hospital.
    - Manufacturing guidance from Actim Partus suggests that the swab should be collected from the cervical os. Visualising the cervix can sometimes be difficult, therefore it may not always be practical to take the sample from the advised cervical os.
- Some of these tests have dual purpose since they can be used for other indications e.g. women with multifetal pregnancies, women with ruptured membranes, women not presenting with symptoms – our population is not representative of the whole population of women presenting with TPL. It is however recognised that treatment of these other populations may involve different management strategies.
- The need to integrate networks of maternity and neonatal services
- The potential implications of adopting new biochemical tests on neonatal unit workload and service planning
- Health care service and travel needs and costs borne by patients and relatives due to changes in in-utero transfer policy resulting from new tests. This includes the effect on the likelihood that a very preterm infant is born in a hospital with inadequate level of specialisation, and the cost and health consequences associated with post-natal transfers
- Effects on resource use in other parts of the public services (e.g. educational services) associated with improvements in preterm birth survival rates and avoidance of neonatal long term morbidity
- Long term parental and societal economic impacts
- Equity implications of changes in hospital admissions across rural and urban areas

8 Discussion

8.1 Review of test accuracy evidence

This is the first review to systematically review the biomarker tests Actim Partus, PartoSure and quantitative fFN (at thresholds other than 50ng/ml) together.

A summary of the key findings from the systematic review of test accuracy evidence can be found in section 2.4, page 90 onwards. In brief, 20 included studies evaluated an index test against the 7-day reference standard [1, 44-64] and seven studies evaluated an index test against a 48hr reference standard.[44, 51, 54, 55, 58-60]

There was only sufficient evidence for pooling the test accuracy data for Actim Partus and PartoSure against the 7-day reference standard and Actim Partus against the 48hr reference standard. However, given that there was substantial methodological, clinical and statistical heterogeneity between studies and large 95% prediction regions (Actim Partus against the 7 day reference standard) and wide 95% confidence intervals (PartoSure against the 7 day reference standard and Actim Partus against the 48hr reference standard), there are considerable uncertainties surrounding the validity of these results.

Studies which offered the greatest certainty when looking to compare test accuracy results were those that assessed two or more different tests within the same population. We identified two such studies: APOSTEL-1 (2016, assessing Actim Partus and quantitative fFN)[45, 46] and Hadzi-Lega (2017, assessing Actim Partus and PartoSure).[47] No studies assessing quantitative fFN and PartoSure within the same population were identified by our review. From APOSTEL-1, the sensitivity was superior for quantitative fFN at the thresholds 10 and 50ng/ml compared to Actim Partus, while Actim Partus had a superior sensitivity against quantitative fFN at the 200ng/ml and 500ng/ml thresholds. Whilst specificity was superior using Actim Partus compared to quantitative fFN at the thresholds of 10, 50 and 200ng/ml but not against the threshold of 500ng/ml. From Hadzi-Lega, the sensitivities were the same between PartoSure and Actim Partus while the specificity was superior using PartoSure. However, the confidence intervals for all sensitivity and specificity data from the two tests assessed within these two studies overlapped considerably (all data available in Table 8 – page 86 or Table 10 – page 97).

When looking at the ranges of results for individual tests across studies, substantial heterogeneity between the studies is clearly apparent. For Actim Partus against the 7 day reference standard (n=16 studies), the study with the best overall sensitivity and specificity results was Tripathi (2016, 94.7% (95%CI 89.9, 97.7) and 92.4% (95%CI 88.9, 95.1))[59] while Cooper (2012) reported the worst sensitivity 33.3% (95%CI 4.3, 77.7) and specificity 74.1%

(95%CI 69.1, 78.6)).[1] These two studies make up two of the three largest studies identified in our review (Tripathi 2016 n=468 and Cooper 2012 n=349). For PartoSure against the 7 day reference standard (n=4 studies), the study with the best overall sensitivity and specificity results was Bolotskikh (2017, 100.0% (95%CI 73.5, 100.0) and 95.4% (95%CI 88.6, 98.7))[61] whilst Werlen (2015) reported the worst (sensitivity 0.0% (95%CI 0.0, 97.5) and specificity 97.5% (95%CI 96.8, 99.9)).[44] The low sensitivity, from Werlen (2015), is attributable to only one woman testing (falsely) positive within the sample of 41. If discounting the results from Werlen (2015), the next study reporting the worst overall sensitivity and specificity results was Nikolova (2015, 80.0% (95%CI 63.1, 91.6) and 94.6% (95%CI 90.1, 97.5)).[62, 63] Fetal Fibronectin at a threshold of 10ng/ml (n=2 studies) had a sensitivity range of 93.8% (95%CI 82.8, 98.7) to 95.7% (95%CI 87.8, 99.1) and a specificity range of 32.2% (95%CI 27.7, 37.0) to 42.3% (95%CI 36.5, 48.4); at a threshold of 200ng/ml had a sensitivity range of 70.8% (95%CI 55.9, 83.0) to 71.0% (95%CI 58.8, 81.3) and a specificity range of 78.6% (95%CI 74.3, 82.5) to 83.6% (95%CI 78.8, 87.8) and at a threshold of 500ng/ml had a sensitivity range of 29.2% (95%CI 17.0, 44.1) to 42.0% (95%CI 30.2, 54.5) and a specificity range of 94.3% (95%CI 91.6, 96.4) to 95.7% (95%CI 92.7, 97.8). Looking at this data, given the large ranges between studies assessing the same test and the wide confidence intervals, it would be premature to attempt to deduce which test was superior against the 7-day reference standard.

We were only able to assess Actim Partus and PartoSure against the 48hr reference standard, since no studies were identified for quantitative fFN. From the single PartoSure study (Werlen 2015) the sensitivity was 0.0% (95%CI 0.0, 97.5) and specificity 97.5% (95%CI 86.8, 99.9); the total sample size was 41 and only one test result was positive (a false positive).[44] From the six Actim Partus studies, the data could be pooled, however the same heterogeneity issues as with studies against the 7-day reference standard were relevant here too. Looking at Actim Partus against the 48hr reference standard (n=6 studies), the study with the best overall sensitivity and specificity results was Tripathi (2016, 95.5% (95%CI 89.7, 98.5) and 82.1% (95%CI 77.8, 86.0))[59] whilst Goyal (2016) reported the worst (sensitivity 65.7% (95%CI 47.8, 80.9) and specificity 56.0% (95%CI 34.9, 75.6)).[54] Given we only identified a single study for PartoSure and the wide range of test accuracy data between the study reporting the best and worst results for Actim Partus, it would also be premature to attempt to deduce which test was superior against the 48hr reference standard.

We identified two relatively recent systematic reviews (Boots 2014 and Sanchez-Ramos 2009) that assessed fFN at a threshold of 50ng/ml, i.e. current practice.[67, 68] Both these reviews suffered from similar heterogeneity issues as our review and their summary ROC

plots displayed large 95% prediction regions. The pooled sensitivities for Boots and Sanchez-Ramos were 75% (95%CI 69, 80) and 76.1% (95%CI 69.1, 81.9) respectively, whilst the specificities were 79% (95%CI 76, 83) and 81.9% (95%CI 78.9, 84.5) respectively.

Both this and the uncertainty about the true accuracy of the index tests inevitably compromises the report's ability to firmly conclude whether the accuracy of the index tests is better than current practice.

8.1.1 Strengths

The strengths of this systematic review are that it was conducted by an independent, experienced research team using the latest evidence and working to a pre-specified protocol (PROSPERO CRD42017072696) that follows a robust methodology.

The search strategy was devised by a dedicated information specialist. The strategy did not restrict by study design and included both forward and backward citation chasing, web-searching and cross-checking with studies provided by the companies. The studies were independently screened by two reviewers, with data extraction and quality appraisal performed by one reviewer and checked by a second.

8.1.2 Weaknesses

The primary weakness of the review of test accuracy was the substantial methodological and clinical heterogeneity between included studies. There was considerable heterogeneity in the following areas: prevalence of preterm birth, mode of delivery, gestational age, definition (symptoms) of preterm labour (including dilation threshold), inclusion of multiple gestations, participant characteristics and provision of treatments. As a consequence, the reported accuracies of individual tests varied widely, hence the confidence intervals in the pooled analyses are also wide. Subsequently, we have limited confidence in the mean pooled test accuracy results.

A limitation to our review, was the lack of published studies where two or more index tests were administered to the same population. Such studies allow us to have more confidence in any differences between the accuracy results of the tests as differences would not be attributable to population or study design. Only two studies APOSTEL-1 (2016)[45, 46] and Hadzi-Lega (2017)[47] assessed the diagnostic test accuracy of two different index tests in the same population; APOSTEL-1 assessed both Actim Partus and quantitative fFN whilst Hadzi-Lega (2017) assessed Actim Partus and PartoSure. We did not identify any studies where all three tests were used in the same population.

Our review was also limited by the comparatively lower number of published studies for quantitative fFN (n=2 studies) and PartoSure (n=4 studies) compared to Actim Partus (n=16

studies) against the 7 day reference standard. In addition, fewer studies published data against the 48hr reference standard with only seven studies being identified (n=6 for Actim Partus and n=1 for PartoSure). Meta-analysis of test accuracy data requires a minimum of four studies. The scope for meta-analysis was therefore restricted. We are aware of three studies published after our searches were run which assess PartoSure (Wing et al. 2017,[140] Lofti et al. 2017,[141] and Melchor et al. 2017 [142]) and [redacted] We also identified seven relevant ongoing trials, four of which are UK based and two plan to enrol over 1000 participants. Personal communications with trial organisers indicate that data from the two large UK trials (QUIDS / QUIDS-2 assessing PartoSure, Actim Partus and quantitative fFN and PETRA assessing quantitative fFN) is expected in 2018. There is the potential, should our analyses be re-run using the data from these trials, the estimates of relative test accuracy may change.

The scope issued by NICE[13] asked for an assessment of test accuracy of quantitative fFN at thresholds other than 50ng/ml. Our capabilities to look at different thresholds was limited by those reported in the published studies. The two quantitative fFN studies both used thresholds of 10, 50, 200 and 500ng/ml. Without access to the individual patient data, we were unable to assess any other thresholds.

Due to the paucity of published test accuracy studies, we made two protocol amendments. The first was to include women with multiple gestations (up to 20% of the total population). Without this amendment there would have been no includable quantitative fFN studies. The second was to include studies where testing was not carried out in-line with clinical practice (i.e. the samples were frozen and analysed at a later date). Without this amendment, we would have had only one includable quantitative fFN study (EUIFS).[64]

Dependent on how the data was reported by each study, we were required to perform some data manipulations. Most studies reported the raw TP, TN, FP and FN data enabling us to calculate additional test accuracy statistics, such as sensitivity, specificity, PPV and NPV. However, four studies (Brik 2010, Riboni 2011, Ting 2007 and Tripathi 2016)[51, 56, 58, 59] only reported sensitivity, specificity PPV and NPV. We back calculated from these statistics to derive the TP, TN, FP and FN values, which were required in our review. All the raw data as reported in the published studies is available in Table 55 and Table 56.

Our review limited included studies to those published in the English language. This may be considered a limitation, however a systematic review assessing the bias of excluding studies that were not in the English language found no evidence of bias.[143] Our review did include a

French study, (Werlen 2015) as a certified translation was received from the manufacturer (Parsagen).

8.1.3 Areas of uncertainty

There is considerable uncertainty surrounding the generalisability of the studies to the UK population. Most specifically none of the included studies were conducted in the UK. In addition, when looking at the prevalence rates of preterm birth of our included studies, prevalence ranged from 2 to 73%. UK prevalence is approximately 8%. These differences in prevalence between studies are likely due to the differences in the populations recruited into the studies (e.g. differences in gestational age, in presenting symptoms of preterm labour, and in recruitment of high or low risk women). It is likely that the prevalence of preterm birth will impact upon the diagnostic test accuracy data presented in section 2.2.6 (page 75).[144] Indeed, Leeflang 2013 explored how sensitivity and specificity vary with disease prevalence, and suggested using prevalence as a guide when selecting studies that most closely match the situation under assessment.[144]

Our ongoing trial searches identified four relevant ongoing UK trials, two of which are very large (over 1,000 participants) and their results are anticipated to be published in 2018 (PETRA and QUIDS).

There is some uncertainty around whether the studies included in the review will be representative of women who do not (or cannot) have access to cervical length measurement. No studies were identified that were specifically based on such a population: the majority of studies did not mention access to cervical length measurement, seven studies used (but did not select participants based upon) cervical length measurement and two studies only included women with a transvaginal cervical length measurement ≤30mm. It is in these latter two studies (APOSTEL-1, Danti 2011) where the most uncertainty occurs; selection based on cervical length measurement would likely increase the prevalence of preterm birth in these studies.[45, 46, 52] However, given that 15 of the 20 included studies had a prevalence rate less than 25% and both studies had a prevalence less than 25% (APOSTEL-1,19.7% and Danti 2011, 6.7%) it is unlikely this criteria impacted prevalence rates.

There was also uncertainty around whether any management strategies (e.g. treatments) would incorrectly inflate false positive rates. As described in section 2.2.4 (page 64) the types of treatments offered to women in threatened preterm labour differed between each study, as did the level of detail describing what and who received the treatments. More often than not, the treatment options were left to the clinician's discretion. It is most likely that the

tocolytic treatments would have the biggest impact on incorrectly inflating the false positive rate, as their purpose is to delay delivery.

It is understood that clinicians would use the results of these tests in combination with other clinical information to make clinical decisions. A mobile application called QUIPP is used in local clinical practice to assist with decision making. The QUIPP app generates a risk score from the following information: whether the mother is symptomatic, how many foetuses, gestation in weeks and days, quantitative fFN value and/or cervical length measurement. In our review, we did not consider combining test results with such clinical data, since no studies were identified that assessed this combination.

We also acknowledge four recently or imminently published studies on PartoSure and quantitative/qualitative fFN. These studies were published after our searches were ran and consequently were not eligible for the review. The studies are:

- Wing et al. 2017,[140] n=796, PartoSure and fFN at 50ng/ml
- Lofti et al. 2017,[141] n=132, PartoSure
- Melchor et al. 2017,[142] n=420, PartoSure and fFN at 50ng/ml
- [redacted]

8.2 Review of clinical effectiveness evidence

No studies were identified which met the inclusion criteria of this review.

8.2.1 Strengths

The review was conducted by an experienced research team and the conducted searches were very sensitive as no study design filters were used. All citations were screened by at least two members of the review team. The review team worked to a pre-specified prospective protocol.

8.2.2 Weaknesses

The review focused on published literature indexed by bibliographic databases, meaning that grey literature was not identified. The searches were conducted in July 2017, so it is possible that studies have been published and indexed subsequently which have not been identified.

8.2.3 Areas of uncertainty

Since no studies were identified for inclusion, we were unable to assess whether using these biomarker tests for predicting preterm labour is clinically effective, i.e. whether they would improve health outcomes.

8.3 Review of cost-effectiveness evidence

Only one conference abstract was identified that was relevant to our cost-effectiveness review.

8.3.1 Strengths

Our review was able to highlight the major developments in methodological practice. Studies have evolved from evaluating only the cost differences of diagnostic strategies, so that competing options were selected solely on the basis of their ability to rule out cases of unnecessary treatment and admission, to evaluating the neonatal health implications of missing the rare cases of true preterm labour. A clear finding from the review was the limited information that most previous economic analyses provide for guiding decision making. With one exception, the model that informed the 2015 NICE Guidelines on diagnosis and treatment of preterm labour, the previous studies do not account for the gestational age gradient in neonatal mortality and morbidity risk exposure and its consequences for cost-effectiveness.

Further, previous models did not account for the variation in costs and benefits of diagnostic testing across hospital settings.

Our results also highlight the need to account for the neonatal hospital cost implications of saving a preterm infant, which no previous study has addressed.

Our analysis also suggests that since the 2015 NICE guidelines on preterm labour diagnosis and treatment, the evidence on the risk-benefit profile for tocolysis and steroids has changed. Tocolysis is now used sparingly, while the importance of providing steroids within 2 days of preterm delivery has gained consensus, especially following articles published earlier this year (2017). Unlike the NICE guidelines model, which used tocolysis as the mediator of improved diagnosis and clinical effectiveness in terms of neonatal mortality and morbidity, cost-effectiveness analysis should be considered with reference to the timing of corticosteroid administration. Despite the 2015 NICE guideline recommendation that women presenting with symptoms of preterm labour below 30 weeks gestation should be admitted to hospital without testing, current practice has not followed this recommendation. Instead, fFN and Actim Partus tests are commonly used to guide admission and treatment decisions in women at early as 22 weeks' gestation. The qualitative fFN test that previously produced a binary result has now been replaced with a test that provides a concentration level. Clinicians are using the new test in more flexible ways than the older binary test, in some cases applying different thresholds for admission and steroid treatment. This warrants new analysis that takes into account the emerging evidence and updated testing practices.

8.3.2 Weaknesses

We focused on economic studies in symptomatic women with intact membranes, and did not cover studies of asymptomatic women where important relevant evidence on utilities, costs, epidemiological parameters and modelling methods may have been generated. For example, we learned that the only study measuring generic health related quality of life for outcomes of mothers (EQ-5D utilities) was a RCT in asymptomatic women (the OPPITIMUM trial, Norman et al. 2016).[145] Although we contacted the authors for data these were not yet available to external researchers to the trial.

At the time of writing, there is no published full economic evaluation of new index tests. The existing studies have only addressed the question of the cost-effectiveness of testing versus no testing. We are aware that there is a research article being prepared for publication on the basis of the QUIDDS project that evaluates the use of quantitative fFN in 1500 individuals across the UK (Sarah Stock personal communication October 2017). This will constitutes the largest known economic study to date on a biochemical test in the population of interest to this review, and will allow to investigate key outcomes for clinical effectiveness such as the ability to predict women delivery within two days of presentation with symptoms suggestive of preterm labour. Previous diagnostic studies of fFN tests have provided limited data on such outcome.

In terms of methodological economic evaluation practice, older studies evaluating the at the time standard practice of treating all symptomatic women assumed that the economic value of diagnostic testing depended solely on a test's ability to rule out cases that were not in preterm labour or likely to delivery within 7 days. More recent studies have extended the analysis from a purely cost-minimisation to a cost-effectiveness framework, by recognising that however few false negatives cases may be missed by testing they are placed at a risk relative to the treat all alternative management option, so that there is a trade-off between cost savings and increased risks to life and quality of life for a few cases with testing. The model that informed the 2015 NICE guideline on diagnosis and treatment of preterm labour represented a methodological advance in that it explicitly recognised that such trade-offs varied by gestational age and were more favourable to testing at older gestational ages.

Since the 2015 NICE guidelines on the topic new evidence has emerged on the value of timely use of ANS, which may still confer benefits to the neonate if given within 1 day of delivery and have a maximum benefit when given between 1 and 2 days of delivery. Existing models however are ill suited to account for this emerging evidence as few diagnostic studies are large enough to include sufficient numbers to measure such outcomes in a reliably, in statistical terms. Current clinical guidelines and recommendations maintain that steroids also need to be given within 7 days of delivery to be effective although empirical

findings published this year by separate independent groups of researchers suggest that some there may have residual beneficial effects when steroids are given more than 7 days before delivery. On the other hand existing studies have not accounted for the risks posed to the neonate by ANS in terms of birthweight, despite the common perception by health professionals on the importance of this risks as manifested through the existing tendency not to give repeated courses of ANS to women who do not deliver within 7 days of an initial course.

8.3.3 Areas of uncertainty

Our review could not inform the cost-effectiveness of PartoSure, Actim Partus, and fFN. The only evidence available was found for fFN in an unpublished study (Master's dissertation) which suggests that fFN at a threshold of 50 ng/ml for treatment and hospital admission may be an inefficient use of NHS resources and that restricting treatment and admission by raising the threshold to 200 ng/ml may be cost-effective. These findings were derived by the AG from the number needed to test to adequately treat a women with steroids reported by the unpublished study of fFN, using a decision tree model and costs and utilities as used by the NICE 2015 guideline model. In view of the limitations of the unpublished model highlighted above, it is unclear whether these findings are robust to sampling and structural model uncertainty.

Key areas of structural uncertainty include the maintained assumption in the NICE model that false negatives (i.e. those who test negative but deliver within 7 days of testing) miss treatment and therefore are placed at increased risk of neonatal death and experiencing adverse chronic events including respiratory distress syndrome and intraventricular haemorrhage. In fact some of those 'missed' cases are likely to return to the maternity hospital and receive ANS closer to delivery thus paradoxically deriving more benefit from the treatment than if they had been detected in the first place. Another key unknown is the effect of accounting for neonatal costs on the results, since previous studies have ignored the costs associated with neonatal deaths and have used data on costs of neonatal morbidity of low quality. Better data on neonatal hospital costs (length of stay at different levels of care) are available for the UK from the National Neonatal Research Database.

8.4 Independent economic assessment

In order to address some of the key limitations in the evidence based, primarily the lack of any evidence on the cost-effectiveness of the index tests in question, we developed a de novo model. The model incorporate the main elements of existing published models, where a decision tree is used to evaluate the costs of the diagnostic phase until delivery, and is linked to data on neonatal outcomes and hospital costs that are mediated by antenatal

steroid use, which is in turn contingent on diagnostic test results. Following the practice in the model that informed the 2015 NICE guideline on the topic, we extrapolated costs and quality adjusted life years to the lifetime of the infant to account for the lasting cost and health related quality of life consequences of neonatal IVH, and the quality of life of infants who survive the first year after birth. Our analysis compared all index tests

- PartoSure
- Actim Partus
- Quantitative fFN at thresholds other than 50 ng/ml

with the comparators:

- fFN 50 ng/ml
- no testing and treating all,

based on the best available diagnostic accuracy evidence for the tests. This turned out to be the two available studies that compared at least two index tests in the same patient sample (Bruijn et al. 2013, Hadzi-Lega et al. 2017). One study compared quantitative fFN with Actim Partus in a group of women from the Netherlands (n=350), while the other compared Actim Partus with PartoSure in a group of women from Macedonia (n=37). We thus presented an economic evaluation of those two comparisons and also of the indirect comparison of PartoSure vs. fFN via Actim Partus as the common treatment to both studies, and included the no testing treat-all and fFN 50 ng/ml comparators using the data from the Dutch study.

In women presenting at 30 weeks' gestation, PartoSure was cheaper and had the same effectiveness as Actim Partus, which in turn saved costs ng/ml at the expense of inferior health outcomes (lower QALYs) relative to fFN 50 ng/ml. 'No testing treat all' and fFN 10 ng/ml had incremental cost-effectiveness ratios of above £100,000 relative to fFN 50 ng/ml, while fFN 200 ng/ml and fFN 500 ng/ml each produced less cost savings per unit of QALY lost relative to fFN 50ng/ml than Actim Partus. Probabilistic Sensitivity Analysis shows that the willingness to pay per QALY would need to be above £100,000 for a test other than PartoSure to become the option most likely to be cost-effective. Similar results apply to women presenting at different gestational ages.

It must be noted that these results are based on two studies from non-UK populations. This is important because diagnostic accuracy results may vary with the prevalence rate and given that the only UK study (n=299; Abbott et al. 2013) reported a 7 day prevalence rate of 3% vs. 19.7% (n=350; APOSTEL-1, Bruijn et al. 2016) and 10.5% (n=57; Hadzi-Lega et al. 2017) reported by the studies used in the base case economic analysis. Subject to this strong caveat, our result that Actim Partus saves more costs per QALY lost relative to fFN

50 ng/ml than qfFN at 200ng/ml and 500ng/ml is robust to different assumptions. The only scenario when this result did not apply was when we used data from a Canadian study (Cooper et al. 2012), however, this appears to have used a qualitative fFN test technology that is no longer in use.

Contrary to the results reported by the economic analysis that informed the 2015 NICE guidelines on preterm labour diagnosis and treatment, we found that the policy of not testing and treating all women presenting at 30 or less weeks of gestation had ICERs well above the £20,000 per QALY gained level. There are important differences between our model and that used by NICE, primarily in terms of the test accuracy data used, which were not available to NICE at the time, and the fact that in our model the mechanism from test results to clinical outcomes operated through the use of ANS; NICE assumed that the benefits occurred through the use of tocolysis and populated its treatment effectiveness parameters from neonatal outcomes reported by RCTs of tocolytics, whereas we populated treatment effectiveness with the latest evidence on steroids effectiveness. Other differences occurred in terms of the measured costs, since we included the costs to the NHS generated by infants saved by ANS, which NICE assumed to be zero. Although we used the same source of national statistics on neonatal and adverse event mortality data, we used more recent data than NICE. On the other hand we adjusted the baseline risks values derived from those data for the fact that ANS is now highly prevalent with 83% of preterm infants born in the UK being given at least one ANS dose; the model used by NICE did not adjust for such prevalent use of ANS and thus may have underestimated the QALY benefits from treatment intensive options.

8.4.1 Strengths

We provide new evidence on the cost-effectiveness of new and existing diagnostic tests in use in the NHS. We model the costs and benefits of diagnostic testing on the basis of recent evidence on the optimal time to delivery intervals for effective antenatal corticosteroid administration in terms of neonatal mortality and morbidity (RDS and IVH). Furthermore, adopt the best modelling practice in the field and we introduce some innovations by accounting for

- gestational age,
- hospital setting (level 1, 2, or 3),
- costs of in utero transfers at very preterm gestational ages,
- neonatal hospital costs of saving a preterm infant,

- Long term QALYs and costs of additional preterm birth survivors with and without adverse events (RDS and IVH)
- Mothers QALYs (in exploratory analysis)

We also conducted extensive scenario analysis and probabilistic analysis to reflect sampling uncertainty in model parameter values.

8.4.2 Weaknesses

We were unable to consider multi-purpose uses of some of the tests (i.e. relevant for other indications). In our analysis we did not account for the effect of costs of any deals offered by suppliers for the purchasing of combinations of diagnostic tests or in bulk over a hospital network.

The scope of our analysis is limited by the fact that we do not consider diagnostic strategies involving combination of tests. The population was defined as women for whom transvaginal ultrasound was not indicated or attending maternity hospitals where it was unavailable.

We had to approximate the neonatal hospital costs of RDS and IVH, assuming they are the same and equal to the additional neonatal hospital costs incurred by infants who were admitted to BAPM level 1 care (this led to costs similar to those used by the NICE 2015 Guideline model).

Critically, the evidence on accuracy was limited to two comparative studies, one of which may be considered small (n=57) to reliably detect differences between index tests; in fact the study found PartoSure and Actim Partus to have the same sensitivity, which in our model determines clinical effectiveness. Therefore our findings must be considered with caution and point to the need to conduct further research before drawing any conclusions on the relative cost-effectiveness between Actim Partus and PartoSure. On firmer ground are the findings that, relative to fFN 50 ng/ml, Actim Partus produced larger savings per QALY lost than fFN 200 ng/ml and fFN 500 ng/ml, whilst 'treat all' and fFN 10 ng/ml had ICERs above the £80,000 mark in women presenting at 24-34 weeks; these results are based on diagnostic accuracy data from one study (Bruijn et al, 2013) in 350 women.

It is worth not in that the only UK diagnostic study was a non-comparative study of quantitative fFN, in which fFN 200 ng/ml and 500 ng/ml resulted in the same sensitivity to that of fFN 50 ng/ml. This was considered in scenario analyses, and fully incremental analysis suggests that lowering the fFN threshold for diagnosing preterm labour from 500 ng/ml to 200 ng/ml has an ICER below £20,000 whereas lowering it from 200 to 50 ng/ml has an ICER above £20,000.

There is an absolute lack of evidence on the outcomes of the status quo diagnostic test, quantitative fFN, in routine practice. Our analysis was largely based on the modelling of the strict adherence to local hospitals' treatment guidelines with fFN. We were able to obtain audit data on treatment practice over six months (2016-2017) in a local level 2 hospital (n=75); however, this is too small an administrative sample to draw definitive conclusions.

A methodological limitation of our analysis is the working assumption that false negative cases miss treatment altogether with the associated increased exposure to risks of neonatal mortality and morbidity. It may be argued that at least some false negatives may actually end up benefitting from being 'missed', since they may deliver between 2 and 7 days, and therefore unintendedly get the chance to return before delivery and closer to the target optimal window for ANS administration of 1 to 2 days before delivery (Norman et al. 2017). Thus many if not most false negatives may possibly end up having the same clinical outcomes as true negatives. This would appear to justify applying a simpler modelling approach that focused only on costs (i.e. cost-minimisation analysis), or perhaps on 48-hour as opposed to 7-day diagnostic outcomes. Given the limited existing data and large sample sizes required to measure 48-hour diagnostic outcomes, the cost-minimisation analysis may seem the only practicable alternative and we conducted scenario analyses that limited the analytical horizon to the time of delivery, effectively providing a cost-minimisation analysis of the decision problem of interest (i.e. since we do not measure any clinical outcomes for the diagnostic phase). This scenario analysis also favoured Actim Partus and suggested that PartoSure may be cost-effective subject to the strong caveat discussed above. Another argument in support of the cost-minimisation is provided by the perceived harm of multiple course of ANS; in this case the key diagnostic outcome measure for clinical effectiveness would be the false positive rate since it would result in not only ineffective use of ANS to start with but it would measure the extent to which women were precluded from any use of ANS at thus exposed to the risk of adverse neonatal health outcomes. This suggests the cost minimisation analysis may be thought of as a conservative alternative scenario to the base case.

We could found no reliable data to account for the side effects of ANS in terms of birthweight. The only source of data that we identified on long term quality of life outcomes by birthweight was from a Canadian study by Saigal and colleagues,[116] which involved measured quality of life outcomes in 286 extremely low birthweight survivors at adolescence and young and mature adulthood. Our analysis that data produced no detectable relationship between birthweight and quality of life, which may be attributed to the sample being too small to reliably measure long term outcome differences by gestational age.

Also, there was little evidence on quality of life outcomes of mothers, and could only perform some exploratory analysis of results including such evidence. Another limitation is that costs and utility values were independent from gestational age at birth; for example we assumed that provided a preterm infant survives until the first birthday, they will get experience the same long term quality of life as the average population. Due to lack of data on the quality of life effects of IVH we had to use utility values from other patient populations, i.e. individual with intracranial haemorrhage.

8.4.3 Areas of uncertainty

There is a high degree of uncertainty associated with the costs and health benefits of PartoSure relative to other tests, as we only had access to one comparative study of this study which involved less than 100 individuals. A study of this size is unlikely to provide reliable results on diagnostic accuracy on delivery outcomes at 48 hours and 7 days. Uncertainty of our findings is compounded by the lack of a study of all relevant tests in the same patient sample, which led us to resort to indirect comparisons using common comparator (i.e. Actim Partus) approach.

More generally as discussed above the ability to predict delivery within 48 hours has become critical, given recent evidence on the importance for good neonatal outcomes of administering ANS within such a short period before birth. In addition to prediction within 7 days, we aimed to differentiate in our model for the ability of a test to predict delivery within 48 hours but we abandoned that analysis due to lack of the required data for the great majority of studies. Further diagnostic test accuracy studies in preterm labour should be undertaken with sufficiently large samples to measure and report diagnostic outcomes for the within 48-hours delivery end point in addition to the conventional 7-day outcomes.

Given the importance of this element for costs-effectiveness of diagnostic options, further research is warranted to produce the data required to calculate more precise measures of long term QALYs gained by saving a neonatal life, which distinguishes by the gestational age at birth. Furthermore, evidence is required on the long term implications of birthweight reductions on quality of life, so that an adequate picture of the implications of the benefit-risk profile of ANS may be accounted for in economic assessments.

There is a high degree of heterogeneity in the evidence on diagnostic test accuracy, as illustrated by the very different prevalence of preterm birth in the largest comparative study (APOSTEL-1, 20% delivery within 7 days) and the UK study (Abbott et al. 2013, 3%), which is not explained by the former study's inclusion and the latter's exclusion of non-spontaneous preterm births. More obvious differences in selection of patients for study participation is also a problem as highlighted by the wide variation in preterm birth

prevalence between one of the four available studies comparing Actim Partus and fFN 50 ng/ml, which (Tripathi et al. 2016), by including patients at gestational ages of 28-36, rather than the more common 24-34 range, led to a 7-day prevalence of preterm birth of 32.5 vs. 19.7% in the largest comparative study of these tests (APOSTEL-1) (see section 2.2.2.1.3).

Some of the uncertainties discussed above may be addressed by the findings of two large (over 1000 participants) ongoing trials of the predictive utility of quantitative fFN (PETRA) and quantitative fFN, PartoSure and Actim Partus (QUIDS 2) in the UK. Until these new data becomes available, the value of PartoSure is unlikely to be settled.

8.5 Other relevant factors

In these analyses we have not allowed for the complicated treatment protocols and guidelines that have been implemented in some hospitals, whereby, for example, some women may be admitted with a concentration level above one fFN threshold but not given antenatal steroids unless that level is above another, higher threshold. This type of more complicated decision algorithm may work to optimise the status quo, but may more reasonably be evaluated using actual data on its operation and results rather than, as we have done here, modelled on the assumption of full adherence.

9 Conclusions

Whilst evidence was identified relating to the test accuracy of the three biomarker tests, there was too much uncertainty in the results to be able to draw any clear conclusions on the relative accuracies of the tests. With the imminent publication of relevant UK based data from large studies, it may be advisable to wait for this data to be published.

We identified no trials that followed patients from testing to ultimate health outcomes. Therefore we found no direct evidence on whether the use of the biomarker tests for predicting preterm labour lead to improved health outcomes. Instead, we used modelling for this purpose.

The limited evidence from diagnostic test accuracy studies in patient populations from non-UK countries suggest that Actim Partus may result in a QALY loss of 0.006 per women tested but reduce health care costs by more than £30,000 per QALY lost, which is larger than savings achieved by other options that restrict treatment relative to standard practice (qfFN 50 ng/ml), i.e. qfFN at 200ng/ml and qfFN 500ng/ml. Also, options that increase access to treatment, i.e. qfFN at 10ng/ml and the policy of no testing and treat all have ICERs in excess of £100,000. As for PartoSure, the evidence is inconclusive due to the small number of patients in the only comparative study of the test. These findings warrant reconsideration in the light of forthcoming evidence from large ongoing diagnostic studies of these tests in UK populations. Our results suggest that the current NICE recommended policy of treating all symptomatic women presenting at less than 30 weeks gestation without testing may not be cost-effective.

9.1 Suggested research priorities

The primary research priority would be for diagnostic test accuracy studies to assess more than one of the index tests within the same trial. Given the practical limitations of comparative studies in this area, a feasible study would give all new tests to the same mothers and compare diagnostic accuracy results with those of the local standard practice. This would allow for a more robust comparison between tests, since population differences between trials would not be an issue. It is probable that these types of studies are currently underway, since some of the index tests are comparatively in their infancy of use and to date are compared predominately to older (qualitative fFN) or not commonly used (cervical length) tests.

New diagnostic studies involving larger samples are required to investigate the differences in available tests in terms of test accuracy outcomes defined relative to delivery within 48 days of testing. Few previous studies have reported these outcomes, likely as a reflection of the

fact that they have included very small samples to reliably measure such outcome. The importance of this question is highlighted by the emerging evidence documenting the first 48 hours as the target window to optimise ANS treatment.

Also required are studies that follow mothers and babies from testing through to final health outcomes. Ideally, such studies would also compare tests. The available evidence only includes observational before-and-after studies of changes in local practice from a policy of managing women according to no testing (treat all) to one based on the results of fFN 50 ng/ml testing. Similar evidence may be obtained by taking advantage of variation in practice across the county, where one of fFN or Actim Partus may be used routinely.

More evidence is required about the side-effects of ANS treatment. Despite the perceived risks of steroids in terms of neonatal outcomes, there is little evidence on the effects of inappropriate use of the treatment.

There is practically no evidence on the mothers' health related quality of life outcomes after diagnostic testing for preterm labour, both before and after birth and over the long term. Observational studies may be able to provide some of these data, particularly in relation to long term outcomes, using existing representative surveys of birth cohorts.

Also warranted would be to improve on the costs of cognitive, respiratory and intestinal neonatal adverse events, using electronic records from the Badger, such as that extracted by the National Neonatal Research Database. This would allow to assess more precisely the costs to the NHS from key neonatal outcomes affected by diagnostic testing.

Acknowledgements

We are grateful for the administrative assistance of Sue Whiffin and Jenny Lowe (including document retrieval). We thank Sophie Robinson for her assistance with running additional searches.

Contributions of authors

Jo Varley-Campbell	Led and contributed to all aspects of the protocol and systematic review of diagnostic test accuracy and clinical effectiveness. Assisted with the systematic review of economic models and drafting of the background. Provided overall project management. Contributed to the writing and editing of the report.
Rubén Mújica-Mota	Contributed to designing the protocol, led the economic evaluation, developed the model and conducted the review of cost-effectiveness studies. Contributed to the writing and editing of the report.
Helen Coelho	Contributed to all aspects of the systematic review of diagnostic test accuracy and clinical effectiveness. Contributed to the writing and editing of the report.
Neel Ocean	Conducted the review of health state utility studies, and extracted relevant parameters. Contributed to model development and checking, and to the writing and editing of the report.
Max Barnish	Contributed to all aspects of the systematic review of diagnostic test accuracy and clinical effectiveness. Led the statistical synthesis and meta-analysis. Contributed to the writing and editing of the report.
David Packman	Conducted a review on the cost of adverse events and extracted relevant data. Contributed to model development and to the writing and editing of the report.
Sophie Dodman	Contributed to all aspects of the systematic reviews. Drafted the background section. Assisted with the identification of parameter values for the economic evaluation. Contributed to the writing and editing of the report.
Chris Cooper	Contributed to designing the protocol. Developed and conducted the literature searches for the systematic reviews of diagnostic test accuracy, economic evaluations and review of utilities. Contributed to the writing and editing of the report.
Tristan Snowsill	Contributed to the development of the protocol and to the meta-analysis of diagnostic accuracy results. Contributed to the writing and editing of the report.
Tracey Kay	Advised on current clinical practice and scientific understanding of preterm labour. Provided IPD data collected at the RD&E Hospital. Contributed to the writing and editing of the report.
Neil Liversedge	Advised on current clinical practice and scientific understanding of preterm labour. Contributed to the writing and editing of the report.

Michelle Parr	Advised on current clinical practice and scientific understanding of preterm labour. Contributed to the writing and editing of the report.
Lisa Knight	Advised on current clinical practice and scientific understanding of preterm labour. Provided IPD data collected at the RD&E Hospital. Contributed to the writing and editing of the report.
Chris Hyde	Project director up to June 2017. Provided diagnostic test accuracy advice to both the systematic review and economic evaluation. Contributed to the writing and editing of the report.
Andrew Shennan OBE	Advised on current clinical practice and scientific understanding of preterm labour. Provided IPD data from previous research. Contributed to the writing and editing of the report.
Martin Hoyle	Project director from June 2017 to completion. Contributed to model checking and advised on model construction. Contributed to the writing and editing of the report.

References

1. Cooper S, Lange I, Wood S, Tang S, Miller L, Ross S. Diagnostic accuracy of rapid phIGFBP-I assay for predicting preterm labor in symptomatic patients. J Perinatol. 2012;32(6):460-5.

2. World Health Organization. Preterm Birth [updated November 2016]. Available from: http://www.who.int/mediacentre/factsheets/fs363/en/.

3. Beck S, Wojdyla D, Say L, Betran AP, Merialdi M, Requejo JH, et al. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. Bulletin of the World Health Organization. 2010;88:31-8.

4. Behrman RE, Butler, A.S. Mortality and Acute Complications in Preterm Infants. . Washington (DC): National Academies Press (US); 2007.

5. Moutquin JM. Classification and heterogeneity of preterm birth. BJOG. 2003;110 Suppl 20:30-3.

6. Romero R, Dey SK, Fisher SJ. Preterm labor: One syndrome, many causes. Science. 2014;345(6198):760-5.

7. Goldenberg RL, Culhane JF, Iams JD, Romero R. Preterm birth 1 - Epidemiology and causes of preterm birth. Lancet. 2008;371(9606):75-84.

8. Halimi Asl AA, Safari S, Parvareshi Hamrah M. Epidemiology and Related Risk Factors of Preterm Labor as an obstetrics emergency. Emergency. 2017;5(1):e3.

9. Goffinet F. Primary predictors of preterm labour. Bjog-Int J Obstet Gy. 2005;112:38-47.

10. March of Dimes. Signs and symptoms of preterm labor [updated August 2017; cited October 2017]. Available from: https://www.marchofdimes.org/complications/signs-and-symptoms-of-preterm-labor.aspx.

11. Ross M, Ridout, A., Shennan, A. Promising recent advances in preterm birth management. Frontiers in Women's Health. 2016;1(2):27-33.

12. Peaceman AM, Andrews WW, Thorp JM, Cliver SP, Lukes A, Iams JD, et al. Fetal fibronectin as a predictor of preterm birth in patients with symptoms: A multicenter trial. Am J Obstet Gynecol. 1997;177(1):13-8.

13. NICE. DAP40: Biomarker tests to help diagnose preterm labour in women with intact membranes - Final Scope. 2017.

14. Office for National Statistics. Birth characteristics in England and Wales: 2016. London: Office for National Statistics, 2017.

15. Smith LK, Draper ES, Manktelow BN, Dorling JS, Field DJ. Socioeconomic inequalities in very preterm birth rates. Archives of disease in childhood Fetal and neonatal edition. 2007;92(1):F11-4.

16. Field D, Boyle E, Draper E, Evans A, Johnson S, Khan K, et al. Towards reducing variations in infant mortality and morbidity: a population-based approach. Programme Grants for Applied Research. Southampton (UK)2016.

17. Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller AB, et al. Born too soon: the global epidemiology of 15 million preterm births. Reprod Health. 2013;10 Suppl 1:S2.

18. World Health Organization. Preterm birth. 2016. http://www.who.int/mediacentre/factsheets/fs363/en/

19. Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. The Lancet. 2012;379(9832):2151-61.

20. Doyle LW, Ford G, Davis N. Health and hospitalistions after discharge in extremely low birth weight infants. Seminars in neonatology : SN. 2003;8(2):137-45.

21. Wood NS, Marlow N, Costeloe K, Gibson AT, Wilkinson AR. Neurologic and developmental disability after extremely preterm birth. EPICure Study Group. N Engl J Med. 2000;343(6):378-84.

22. Saigal S, Doyle LW. Preterm birth 3 - An overview of mortality and sequelae of preterm birth from infancy to adulthood. Lancet. 2008;371(9608):261-9.

23. Gregory KE, Deforge CE, Natale KM, Phillips M, Van Marter LJ. Necrotizing enterocolitis in the premature infant: neonatal nursing assessment, disease pathogenesis, and clinical presentation. Adv Neonatal Care. 2011;11(3):155-64; quiz 65-6.

24. Melville JM, Moss TJ. The immune consequences of preterm birth. Front Neurosci. 2013;7:79.

25. Hodek JM, Mittendorf T, Von der Schulenburg MG. Measuring Economic Consequences of Preterm Birth - Methodological Recommendations for the Evaluation of Personal Burden on Children and Their Caregivers. Acta Paediatr. 2010;99:98-.

26. NICE. Preterm labour and birth (full guideline), NICE guideline [NG25]. 2015.

27. Lee SM, Romero R, Park JW, Kim SM, Park C-W, Korzeniewski SJ, et al. The Clinical Significance Of A Positive Amnisure Test(™) In Women With Preterm Labor And Intact Membranes. The Journal of Maternal-Fetal and Neonatal Medicine. 2012;25(9):1690-8.

28. Alere. Alere Actim Partus: A Summary of Clinical Evidence 2012.

29. Alere. Alere Actim Partus Test: Instructions for use2015.

30. Parsagen Diagnostics Inc. PartoSure, Assess the risk of preterm birth: Instructions for use 2015.

31. Hologic. Rapid fFN 10Q Cassette Kit: Instructions for use 2016.

32. Hologic. QuikCheck fFN Sheet.

33. Haas DM, Caldwell DM, Kirkpatrick P, McIntosh JJ, Welton NJ. Tocolytic therapy for preterm delivery: systematic review and network meta-analysis. Bmj-Brit Med J. 2012;345.

34. Elliott JP, Morrison JC, Bofill JA. Risks and Benefits of Magnesium Sulfate Tocolysis in Preterm Labor (PTL). AIMS Public Health. 2016;3(2):348-56.

35. Roberts D, Dalziel S. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. Cochrane Db Syst Rev. 2006(3).

36. Norman M, Piedvache A, Borch K, Huusom LD, Bonamy AE, Howell EA, et al. Association of Short Antenatal Corticosteroid Administration-to-Birth Intervals With Survival and Morbidity Among Very Preterm Infants: Results From the EPICE Cohort. JAMA Pediatr. 2017;171(7):678-86.

37. Roberts D, Brown J, Medley N, Dalziel SR. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. Cochrane Db Syst Rev. 2017(3).

38. Crowther CA, Doyle LW, Anderson P, Harding JE, Haslam RR, Hiller JE, et al. Repeat dose(S) of prenatal corticosteroids for women at risk of preterm birth: Early school-age outcomes (6 TO 8 YEARS') for children in the actords trial. Journal of Paediatrics and Child Health. 2011;47:52-3.

39. Nelson KB, Grether JK. Can magnesium sulfate reduce the risk of cerebral palsy in very low birthweight infants? Pediatrics. 1995;95(2):263-9.

40. NHS Centre for Reviews and Dissemination. Systematic reviews : CRD's guidance for undertaking reviews in health care [3rd ed.]. York: Centre for Reviews and Dissemination, 2009.

41. Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. Journal of clinical epidemiology. 2011;64(6):602-7.

42. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011;155(8):529-36.

43. Deeks J, Bossuyt P, Gatsonis C. Cochrane handbook for systematic reviews of diagnostic test accuracy. Version 10 London: The Cochrane Collaboration. 2010.

44. Werlen S, Raia T, Di Bartolomeo A, Chauleur C. Preterm labor: Reproducibility of detection test of PAMG-1 before and after digital examination, and transvaginal ultrasound cervical length. Gynecol Obstet Ferti. 2015;43(10):640-5.

45. Bruijn M, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM, et al. Quantitative fetal fibronectin testing in combination with cervical length measurement in the prediction of spontaneous preterm delivery in symptomatic women. BJOG: An International Journal of Obstetrics & Gynaecology. 2016;123(12):1965-71.

46. Bruijn MM, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM, et al. Comparison of the Actim Partus test and the fetal fibronectin test in the prediction of spontaneous preterm birth in symptomatic women undergoing cervical length measurement. Eur J Obstet Gynecol Reprod Biol. 2016;206:220-4.

47. Hadzi-Lega M MJ, Helmer H, Hellmeyer L, Markova AD, Poposka A. Comparison of PAMG-1 and phIGFBP-1 Tests for the Prediction of Preterm Delivery in Patients with Preterm Labor. . Open J Obstet Gynecol. 2017;07(03):11.

48. Abo El-Ezz AE, Askar AE. Predictive value of phosphorylated insulin-like growth factor binding protein-1 (PIGFBP-1) (bedside test) in preterm labor. J Egypt Soc Parasitol. 2014;44(2):525-30.

49. Altinkaya O, Gungor T, Ozat M, Danisman N, Mollamahmutoglu L. Cervical phosphorylated insulin-like growth factor binding protein-1 in prediction of preterm delivery. Arch Gynecol Obstet. 2009;279(3):279-83.

50. Azlin MI, Bang HK, An LJ, Mohamad SN, Mansor NA, Yee BS, et al. Role of phIGFBP-1 and ultrasound cervical length in predicting pre-term labour. Journal of Obstetrics & Gynaecology. 2010;30(5):456-9.

51. Brik M HA, Pedraz CC, Perales A. . Phosphorylated insulin-like growth factor binding protein-1 and cervical measurement in women with threatening preterm birth. Acta Obstet Gynecol Scand. 2010;89(2):268-74.

52. Danti L, Prefumo F, Lojacono A, Corini S, Testori A, Frusca T. The combination of short cervical length and phIGFBP-1 in the prediction of preterm delivery in symptomatic women. Journal of maternal-fetal & neonatal medicine [Internet]. 2011; 24(10):[1262-6 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/086/CN-00890086/frame.html.

53. Eroglu D, Yanik F, Oktem M, Zeyneloglu HB, Kuscu E. Prediction of preterm delivery among women with threatened preterm labor. Gynecologic and Obstetric Investigation. 2007;64(2):109-16.

54. Goyal M, Kriplani A, Kachhawa G, Badiger S. Prediction of preterm labor by a rapid bedside test detecting phosphorylated insulin-like growth factor-binding protein 1 in cervical secretions. Int J Gynaecol Obstet. 2016;134(2):165-8.

55. Lembet A, Eroglu D, Ergin T, Kuscu E, Zeyneloglu H, Batioglu S, et al. New rapid bed-side test to predict preterm delivery: phosphorylated insulin-like growth factor binding protein-1 in cervical secretions. Acta Obstet Gynecol Scand. 2002;81(8):706-12.

56. Riboni F, Vitulo A, Dell'avanzo M, Plebani M, Battagliarin G, Paternoster D. Biochemical markers predicting pre-term delivery in symptomatic patients: phosphorylated insulin-like growth factor binding protein-1 and fetal fibronectin. Arch Gynecol Obstet. 2011;284(6):1325-9.

57. Tanir HM, Sener T, Yildiz Z. Cervical phosphorylated insulin-like growth factor binding protein-1 for the prediction of preterm delivery in symptomatic cases with intact membranes. Journal of Obstetrics and Gynaecology Research. 2009;35(1):66-72.

58. Ting H-S, Chin P-S, Yeo GS, Kwek K. Comparison of bedside test kits for prediction of preterm delivery: Phosphorylated insulin-like growth factor binding protein-1 (pIGFBP-1) test and fetal fibronectin test. Annals Academy of Medicine Singapore. 2007;36(6):399-402.

59. Tripathi R, Tyagi S, Mala YM, Singh N, Pandey NB, Yadav P. Comparison of rapid bedside tests for phosphorylated insulin-like growth factor-binding protein 1 and fetal fibronectin to predict preterm birth. Int J Gynaecol Obstet. 2016;135(1):47-50.

60. Vishwekar PS CA, Turakhia N. Prediction of preterm delivery with a novel bedside test. Int J Reprod Contracept Obstet Gynecol. 2017;6(8):3366-71.

61. Bolotskikh V, Borisova V. Combined value of placental alpha microglobulin-1 detection and cervical length via transvaginal ultrasound in the diagnosis of preterm labor in symptomatic patients. Journal of Obstetrics and Gynaecology Research. 2017.

62. Nikolova T, Bayev O, Nikolova N, Di Renzo GC. Evaluation of a novel placental alpha microglobulin-1 (PAMG-1) test to predict spontaneous preterm delivery. J Perinat Med. 2014;42(4):473-7.

63. Nikolova T, Bayev O, Nikolova N, Di Renzo GC. Comparison of a novel test for placental alpha microglobulin-1 with fetal fibronectin and cervical length measurement for the prediction of imminent spontaneous preterm delivery in patients with threatened preterm labor. J Perinat Med. 2015;43(4):395-402.

64. Bruijn MM, Kamphuis EI, Hoesli IM, Martinez de Tejada B, Loccufier AR, Kuhnert M, et al. The predictive value of quantitative fibronectin testing in combination with cervical length measurement in symptomatic women. Am J Obstet Gynecol. 2016;215(6):793.e1-.e8.

65. Kuhrt K, Unwin C, Hezelgrave N, Seed P, Shennan A. Endocervical and high vaginal quantitative fetal fibronectin in predicting preterm birth. Journal of Maternal-Fetal & Neonatal Medicine. 2014;27(15):1576-9.

66. Adeza B. Fetal Fibronectin Enzyme Immunoassay and Rapid fFN for the TLIIQ system: Information for Healthcare Providers.Published: December 2003 Date Accessed: 27 October 2017.

67. Sanchez-Ramos L DI, Zamora J, Kaunitz AM. Fetal fibronectin as a short-term predictor of preterm birth in symptomatic patients. Obstet Gynecol. 2009;114:631-40.

68. Boots AB S-RL, Bowers DM, Kaunitz AM, Zamora J, Schlattmann P. The short-term prediction of preterm birth: a systematic review and diagnostic metaanalysis. AJOG. 2014;210:1-10.

69. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2013;16(2):231-50.

70. Gibson S, Hezelgrave NL, Shennan AH. The impact of quantitative fetal fibronectin on the clinical management of women symptomatic of preterm labour. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2014;99:A151-A5.

71. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Plos Med. 2009;6(7):e1000097.

72. Abenhaim HA, Morin L, Benjamin A. Does availability of fetal firbonectin testing in the management of threatened preterm labour affect the utilization of hospital resources? Journal of Obstetrics & Gynaecology Canada: JOGC. 2005;27(7):689-94.

73. Joffe GM, Jacques D, Bemis-Heys R, Burton R, Skram B, Shelburne P. Impact of the fetal fibronectin assay on admissions for preterm labor. Am J Obstet Gynecol. 1999;180(3):581-6.

74. Berghella V, Saccone G. Fetal fibronectin testing for prevention of preterm birth in singleton pregnancies with threatened preterm labor: a systematic review and metaanalysis of randomized controlled trials. Am J Obstet Gynecol. 2016;215(4):431-8.

75. Dutta D, Norman JE. Pilot study into the efficacy of foetal fibronectin testing in minimising hospital admissions in women presenting with symptoms of preterm labour: a randomised controlled trial of obstetric and neonatal outcomes. Arch Gynecol Obstet. 2011;284(3):559-65.

76. Lee GT, Burwick R, Zork N, Kjos S. Does the use of fetal fibronectin in an algorithm for preterm labor reduce triage evaluation times? The journal of maternal-fetal & neonatal medicine. 2013;26(7):706-9.

77. Deshpande S, van Asselt A, Tomini F, Armstrong N, Allen A, Noake C, et al. Rapid fetal fibronectin testing to predict preterm birth in women with symptoms of premature labour: a systematic review and cost analysis. Health Technology Assessment. 2013;17(32):1-138.

78. Honest H, Forbes C, Durée K, Norman G, Duffy S, Tsourapas A, et al. Screening to prevent spontaneous preterm birth: systematic reviews of accuracy and effectiveness literature with economic modelling. Health Technology Assessment. 2009;13(37):1-627.

79. Chuck A, Nguyen T. Post policy implementation review (PPIR) of rapid fetal fibronectin testing for preterm labour in Alberta. Edmonton AB: Institue of Health Economics, Alberta Canada, 2015.

80. Boyd KA, Briggs AH, Fenwick E, Norrie J, Stock S. Power and sample size for cost-effectiveness analysis: fFN neonatal screening. Contemporary Clinical Trials. 2011;32(6):893-901.

81. Mozurkewich EL, Naglie G, Krahn MD, Hayashi RH. Predicting preterm birth: A cost-effectiveness analysis. Am J Obstet Gynecol. 2000;182(6):1589-97.

82. van Baaren GJ, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM, et al. Cost-effectiveness of diagnostic testing strategies including cervical length measurement and fibronectin testing in women with symptoms of preterm labor. Ultrasound in Obstetrics & Gynecology. 2017:31.

83. Sedgwick P. What is a non-inferiority trial? BMJ : British Medical Journal. 2013;347.

84. Hogg M, Penney, G., Carmichael, J. Audit of care provided and outcomes achieved by community maternity units in Scotland 2005. Scottish programme for clinical effectiveness in reproductive health. 2007.

85. Healthcare Commission. Towards better births—a review of maternity services in England. Commission for healthcare audit and inspection. 2008.

86. Kenyon SL, Taylor DJ, Tarnow-Mordi W, Grp OC. Broad-spectrum antibiotics for spontaneous preterm labour: the ORACLE II randomised trial. Lancet. 2001;357(9261):989-94.

87. Gyetvai K, Hannah ME, Hodnett ED, Ohlsson A. Tocolytics for preterm labor: a systematic review. Obstet Gynecol. 1999;94(5 Pt 2):869-77.

88. Moutquin JM, Effer SB, Milner RA, Mohide PT, Sauve RS, Sinclair JC, et al. Treatment of Preterm Labor with the Beta-Adrenergic Agonist Ritodrine. New Engl J Med. 1992;327(5):308-12.

89. Crowley PA. Antenatal corticosteroid therapy: a meta-analysis of the randomized trials, 1972 to 1994. Am J Obstet Gynecol. 1995;173(1):322-35.

90. Roos C, Spaanderman MEA, Schuit E, Bloemenkamp KWM, Bolte AC, Cornette J, et al. Effect of Maintenance Tocolysis With Nifedipine in Threatened Preterm Labor on Perinatal Outcomes A Randomized Controlled Trial. Jama-J Am Med Assoc. 2013;309(1):41-7.

91. Vis JY, van Baaren GJ, Wilms FF, Oudijk MA, Kwee A, Porath MM, et al. Randomized comparison of nifedipine and placebo in fibronectin-negative women with symptoms of preterm labor and a short cervix (APOSTEL-I Trial). American journal of perinatology. 2015;32(5):451-60.

92. van Baaren GJ, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM, et al. Predictive Value of Cervical Length Measurement and Fibronectin Testing in Threatened Preterm Labor. Obstet Gynecol. 2014;123(6):1185-92.

93. Kind P. A comparison of two models for scaling health indicators. International journal of epidemiology. 1982;11(3):271-5.

94. Cahill AG, Odibo AO, Stout MJ, Grobman WA, Macones GA, Caughey AB. Magnesium sulfate therapy for the prevention of cerebral palsy in preterm infants: a decision-analytic and economic analysis. Am J Obstet Gynecol. 2011;205(6):542.e1-7.

95. Kruse M, Michelsen SI, Flachs EM, Bronnum-Hansen H, Madsen M, Uldall P. Lifetime costs of cerebral palsy. Developmental medicine and child neurology. 2009;51(8):622-8.

96. Alvarez MD, Villamil M, Reyes G. Predictive factors in the genesis of intraventricular hemorrhage in premature infants. Puerto Rico health sciences journal. 1994;13(4):251-4.

97. Norberg H, Kowalski J, Maršál K, Norman M. Timing of antenatal corticosteroid administration and survival in extremely preterm infants: a national population-based cohort study. BJOG: An International Journal of Obstetrics & Gynaecology. 2017.

98. Watson H, Rideout A, Shennan AH. Response to: Preterm labour: summary of NICE guidance. BMJ : British Medical Journal. 2015;351:h6283.

99. Ridout A, Watson H, Shennan A. Antenatal corticosteroids in perspective: rationalising current practice. BJOG: An International Journal of Obstetrics & Gynaecology. 2016;123(7):1070-.

100. Shennan A, Carter J. Threatened Preterm Labour: guideline for management. In: Guy's and St Thomas', editor. 2017.

101. NICE. Guide to the methods of technology appraisal 2013. 2013.

102. Abbott DS, Radford SK, Seed PT, Tribe RM, Shennan AH. Evaluation of a quantitative fetal fibronectin test for spontaneous preterm birth in symptomatic women. Am J Obstet Gynecol. 2013;208(2):122.e1-6.

103. Royal College of Paediatrics and Child Health (RCPCH). National Neonatal Audit Programme 2016 on Annual Report on 2015 data. 2016.

104. Travers CP, Clark RH, Spitzer AR, Das A, Garite TJ, Carlo WA. Exposure to any antenatal corticosteroids and outcomes in preterm infants by gestational age: prospective cohort study. BMJ (Clinical research ed). 2017;356.

105. Grant RL. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. BMJ : British Medical Journal. 2014;348.

106. Radic JA, Vincer M, McNeely PD. Outcomes of intraventricular hemorrhage and posthemorrhagic hydrocephalus in a population-based cohort of very preterm infants born to residents of Nova Scotia from 1993 to 2010. Journal of Neurosurgery: Pediatrics. 2015;15(6):580-8.

107. World Health Organization. WHO recommendations on interventions to improve preterm birth outcomes. . Geneva Switzerland: WHO; 2015. p. 98.

108. Ware JE, Sherbourne CD. The Mos 36-Item Short-Form Health Survey (Sf-36) .1. Conceptual-Framework and Item Selection. Medical Care. 1992;30(6):473-83.

109. Brooks R. EuroQol: The current state of play. Health Policy. 1996;37(1):53-72.

110. Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? Health and Quality of Life Outcomes. 2009;7(1):27.

111. Feingold E, Sheir-Neiss G, Melnychuk J, Bachrach S, Paul D. HRQL and severity of brain ultrasound findings in a cohort of adolescents who were born preterm. Journal of Adolescent Health. 2002;31(3):234-9.

112. Beaudoin S, Tremblay GM, Croitoru D, Benedetti A, Landry JS. Healthcare utilization and health-related quality of life of adult survivors of preterm birth complicated by bronchopulmonary dysplasia. Acta Paediatrica, International Journal of Paediatrics. 2013;102(6):607-12.

113. Alemdaroglu I, Kepenek Varol B, Tanriverdi M, Guler S, Iscan A. Comparison of the level of depression and quality of life of mothers who have low birth weight premature infants with or without intracranial/intraventricular hemorrhage. Fizyoterapi Rehabilitasyon. 2015;26(2):S81.

114. Hill PD, Aldag JC. Maternal perceived quality of life following childbirth. JOGNN - Journal of Obstetric, Gynecologic, and Neonatal Nursing. 2007;36(4):328-34.

115. Couto ER, Couto E, Vian B, Gregorio Z, Nomura ML, Zaccaria R, et al. Quality of life, depression and anxiety among pregnant women with previous adverse pregnancy outcomes. Sao Paulo Medical Journal. 2009;127(4):185-9.

116. Saigal S, Ferro MA, Van Lieshout RJ, Schmidt LA, Morrison KM, Boyle MH. Health-Related Quality of Life Trajectories of Extremely Low Birth Weight Survivors into Adulthood. Journal of Pediatrics. 2016;179:68-73.

117. Carroll AE, Downs SM. Improving decision analyses: parent preferences (utility values) for pediatric health outcomes. The Journal of pediatrics. 2009;155(1):21-5. e5.

118. Carroll, Bastek JA, Langmuir H, Kondapalli LA, Pare E, Adamczak JE, et al. Antenatal corticosteroids for late-preterm infants: a decision-analytic and economic analysis. ISRN Obstetrics & Gynecology. 2012;2012:491595.

119. Fawke J, Lum S, Kirkby J, Hennessy E, Marlow N, Rowell V, et al. Lung Function and Respiratory Symptoms at 11 Years in Children Born Extremely Preterm. American Journal of Respiratory and Critical Care Medicine. 2010;182(2):237-45.

120. Cooke RWI. Health, lifestyle, and quality of life for young adults born very preterm. Archives of Disease in Childhood. 2004;89(3):201-6.

121. Batsvik B, Vederhus BJ, Halvorsen T, Wentzel-Larsen T, Graue M, Markestad T. Health-related quality of life may deteriorate from adolescence to young adulthood after extremely preterm birth. Acta Paediatrica, International Journal of Paediatrics. 2015;104(9):948-55.

122. Lund LK, Vik T, Lydersen S, Lohaugen GC, Skranes J, Brubakk AM, et al. Mental health, quality of life and social relations in young adults born with low birth weight. Health & Quality of Life Outcomes. 2012;10:146.

123. Papaioannou D, Brazier J, Paisley S. NICE DSU Technical Support Document 9: the identification, review and synthesis of health state utility values from the literature. University of Shefflield. 2011.

124. Vandenbussche FP, Jong-Potjer D, Lieke C, Stiggelbout AM, Le Cessie S, Keirse MJ. Differences in the valuation of birth outcomes among pregnant women, mothers, and obstetricians. Birth. 1999;26(3):178-83.

125. Dalziel SR, Lim VK, Lambert A, McCarthy D, Parag V, Rodgers A, et al. Psychological functioning and health-related quality of life in adulthood after preterm birth. Developmental medicine and child neurology. 2007;49(8):597-602.

126. Husby IM, Stray KM, Olsen A, Lydersen S, Indredavik MS, Brubakk AM, et al. Long-term follow-up of mental health, health-related quality of life and associations with motor skills in young adults born preterm with very low birth weight. Health & Quality of Life Outcomes. 2016;14:56.

127. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value in Health. 2010;13(5):509-18.

128. Crump C, Sundquist K, Sundquist J, Winkleby MA. Gestational age at birth and mortality in young adulthood. Jama. 2011;306(11):1233-40.

129. Swamy GK, SkjŠrven R. Association of preterm birth with long-term survival, reproduction, and next-generation preterm birth. Jama. 2008;299(12):1429-36.

130. Office for National Statistics. Pregnancy and ethnic factors influencing births and infant mortality. 2015.

131. Pham C, Crowther C. Birth outcomes: utility values that postnatal women, midwives and medical staff express. BJOG: An International Journal of Obstetrics & Gynaecology. 2003;110(2):121-7.

132. Parisaei M, Currie J, O'Gorman N, Morris S, David AL. Implementation of foetal fibronectin testing: Admissions, maternal interventions and costs at 1 year. Journal of Obstetrics and Gynaecology. 2016;36(7):888-92.

133. Curtis L, Burns A. Unit Costs of Health and Social Care 2016. Personal Social Services Research Unit, University of Kent, Canterbury. 2016.

134. Gale C, Morris I. The UK National Neonatal Research Database: using neonatal data for research, quality improvement and more. Archives of disease in childhood - Education & practice edition. 2016;101(4):216-8.

135. Adams E. The New NCCMDS, Neonatal HRGs 2016 and Reference Costs: A Guide for Clinicians. National Casemix Office, NHS England: 2016.

136. Mangham LJ, Petrou S, Doyle LW, Draper ES, Marlow N. The cost of preterm birth throughout childhood in England and Wales. Pediatrics. 2009;123(2):e312-e27.

137. Gale C, Hay A, Philipp C, Khan R, Santhakumaran S, Ratnavel N. In-utero transfer is too difficult: results from a prospective study. Early human development. 2012;88(3):147-50.

138. Landry JS, Croitoru D, Jin Y, Schwartzman K, Benedetti A, Menzies D. Health care utilization by preterm infants with respiratory complications in Quebec. Canadian respiratory journal. 2012;19(4):255-60.

139. Khan KA, Petrou S, Dritsaki M, Johnson SJ, Manktelow B, Draper ES, et al. Economic costs associated with moderate and late preterm birth: a prospective population-based study. BJOG: An International Journal of Obstetrics & Gynaecology. 2015;122(11):1495-505.

140. Wing DA, Haeri S, Silber AC, Roth CK, Weiner CP, Echebiri NC, et al. Placental Alpha Microglobulin-1 Compared With Fetal Fibronectin to Predict Preterm Delivery in Symptomatic Women. Obstet Gynecol. 2017;130(6):1183-91.

141. Lotfi G, Faraz S, Nasir R, Somini S, Abdeldayem RM, Koratkar R, et al. Comparison of the effectiveness of a PAMG-1 test and standard clinical assessment in the prediction of preterm birth and reduction of unnecessary hospital admissions. The journal of maternal-fetal & neonatal medicine. 2017:1-5.

142. Melchor JC, Navas H, Marcos M, Iza A, de Diego M, Rando D, et al. Retrospective cohort study of PAMG-1 and fetal fibronectin test performance in assessing spontaneous preterm birth risk in symptomatic women attending an emergency obstetrical unit. Ultrasound Obstet Gynecol. 2017.

143. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, et al. The Effect of English-Language Restriction on Systematic Review-Based Meta-Analyses: A Systematic Review of Empirical Studies. International Journal of Technology Assessment in Health Care. 2012;28(2):138-44.

144. Leeflang MMG, Rutjes AWS, Reitsma JB, Hooft L, Bossuyt PMM. Variation of a test's sensitivity and specificity with disease prevalence. Can Med Assoc J. 2013;185(11):E537-E44.

145. Norman JE, Marlow N, Messow CM, Shennan A, Bennett PR, Thornton S, et al. Vaginal progesterone prophylaxis for preterm birth (the OPPTIMUM study): a multicentre, randomised, double-blind trial. Lancet. 2016;387(10033):2106-16.

146. Baumgardt M, Bucher HU, Mieth RA, Fauchere JC. Health-related quality of life of former very preterm infants in adulthood. Acta Paediatr. 2012;101(2):e59-63.

147. Berbis J, Einaudi MA, Simeoni MC, Brevaut-Malaty V, Auquier P, d'Ercole C, et al. Quality of life of early school-age French children born preterm: a cohort study. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2012;162(1):38-44.

148. Bianco F, Cremonesi G, Montirosso R, Borgatt R. Are preterm infants likely to have the same quality of life as those born at term? Do preterm infants treated with surfactant have the same respiratory outcome as those not treated? Results from the NEO-ACQUA study. Neonatology. 2011;99(4):371-2.

149. Ketharanathan N, Lee W, De Mol AC. Health-related quality of life, emotional and behavioral problems in mild to moderate prematures at (pre-)school age. Early Human Development. 2011;87(10):705-9.

150. Gray R, Petrou S, Hockley C, Gardner F. Self-reported health status and health-related quality of life of teenagers who were born before 29 weeks' gestational age. Pediatrics. 2007;120(1):e86-93.

151. Petrou S, Abangma G, Johnson S, Wolke D, Marlow N. Costs and health utilities associated with extremely preterm birth: Evidence from the EPICure study. Value in Health. 2009;12(8):1124-34.

152. Roberts G, Burnett AC, Lee KJ, Cheong J, Wood SJ, Anderson PJ, et al. Quality of life and functional outcomes in extremely preterm or extremely low birth weight 18-year-olds born in the post-surfactant-a regional study. Journal of Paediatrics and Child Health. 2013;49:22-3.

153. van Lunenburg A, van der Pal SM, van Dommelen P, van der Pal de Bruin KM, Bennebroek Gravenhorst J, Verrips GHW. Changes in quality of life into adulthood after very preterm birth and/or very low birth weight in the Netherlands. Health and Quality of Life Outcomes. 2013;11(1):51.

154. Verrips E, Vogels T, Saigal S, Wolke D, Meyer R, Hoult L, et al. Health-related quality of life for extremely low birth weight adolescents in Canada, Germany, and the Netherlands. Pediatrics. 2008;122(3):556-61.

155. Wolke D. Born Extremely Low Birth Weight and Health Related Quality of Life into Adulthood. Journal of Pediatrics. 2016;179:11.

156. Wolke D, Chernova J, Eryigit-Madzwamuse S, Samara M, Zwierzynska K, Petrou S. Self and parent perspectives on health-related quality of life of adolescents born very preterm. Journal of Pediatrics. 2013;163(4):1020.

157. Zwicker JG, Harris SR. Quality of life of formerly preterm and very low birth weight infants from preschool age to adulthood: A systematic review. Pediatrics. 2008;121(2):e366-e76.

158. Korvenranta E, Linna M, Rautava L, Andersson S, Gissler M, Hallman M, et al. Hospital costs and quality of life during 4 years after very preterm birth. Archives of Pediatrics and Adolescent Medicine. 2010;164(7):657-63.

159. Lehtonen L, Rautava L, Korvenranta E, Korvenranta H, Peltola M, Hakkinen U. PERFECT preterm infant study. Annals of Medicine. 2011;43(SUPPL. 1):S47-S53.

160. Rautava L, Hakkinen U, Korvenranta E, Andersson S, Gissler M, Hallman M, et al. Health-Related Quality of Life in 5-Year-Old Very Low Birth Weight Infants. Journal of Pediatrics. 2009;155(3):338.

161. Einerson BD, Grobman WA, Miller ES. Cost-effectiveness of risk-based screening for cervical length to prevent preterm birth. Am J Obstet Gynecol. 2016;215(1):100.

162. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-Attribute Preference Functions. PharmacoEconomics. 1995;7(6):503-20.

163. Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. Medical care. 2000:583-637.

164. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Medical care. 2002;40(2):113-28.

165. Furlong W, Feeny D, Torrance G, Goldsmith C, DePauw S, Zhu Z, et al. Multiplicative multi-attribute utility function for the Health Utilities Index Mark 3 (HUI3) system: a technical report. Centre for Health Economics and Policy Analysis (CHEPA), McMaster University, Hamilton, Canada, 1998.

166. Saigal S, Feeney D, Furlong W, Rosenbaum P, Burrows E, Torrance G. Comparison of the health-related quality of life of extremely low birth weight children and a-reference group of children at age eight years. Journal of Pediatrics. 1994;125(3):418-25.

167. Saigal S, Feeny D, Rosenbaum P, Furlong W, Burrows E, Stoskopf B. Self-perceived health status and health-related quality of life of extremely low-birth-weight infants at adolescence. Jama. 1996;276(6):453-9.

168. Saigal S, Rosenbaum P, Stoskopf B, Hoult L, Furlong W, Feeny D, et al. Comprehensive assessment of the health status of extremely low birth weight children at eight years of age: comparison with a reference group. The Journal of pediatrics. 1994;125(3):411-7.

169. Saigal S, Rosenbaum PL, Feeny D, Burrows E, Furlong W, Stoskopf BL, et al. Parental perspectives of the health status and health-related quality of life of teen-aged children who were extremely low birth weight and term controls. Pediatrics. 2000;105(3):569-74.

170. Saigal S, Stoskopf B, Pinelli J, Streiner D, Hoult L, Paneth N, et al. Self-perceived health-related quality of life of former extremely low birth weight infants at young adulthood. Pediatrics. 2006;118(3):1140-8.

171. Coyle SB. Maternal concern, social support, and health-related quality of life across childhood. Research in Nursing and Health. 2011;34(4):297-309.

172. Kim S-H, Kim S-O, Lee S-i, Jo M-W. Deriving a mapping algorithm for converting SF-36 scores to EQ-5D utility score in a Korean population. Health and quality of life outcomes. 2014;12(1):145.

173. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. Health and quality of life outcomes HERC database of mapping studies, Version 50 (Last updated: 16th May 2016). 2013;11(1):151.

174. Kwon J, Kim SW, Ungar WJ, Tsiplova K, Madan J, Petrou S. A systematic review and meta-analysis of childhood health utilities. Medical Decision Making. Forthcoming.

175. Wolke D, Baumann N, Busch B, Bartmann P. Very Preterm Birth and Parents' Quality of Life 27 Years Later. Pediatrics. 2017;140(3).

Appendix 1. Literature search strategies

A1.1 Database searches

Database: MEDLINE

Host: OVID

Data Parameters: Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R)

Date searched: Monday July 24th 2017

Searcher: CC

Search Checked by: JVC

Hits: 916

#	Searches	Results
1	(PartoSure or Parto Sure or PartoSureR or Parto SureR).ti,ab,kw.	3
2	((Placental alpha adj5 test$) or PAMG-1).ti,ab,kw.	44
3	(Actim Partus or Actim PartusR or "insulin-like growth factor-binding protein 1 test" or phIGFBP-1 or (IGFBP-1 adj5 test$)).ti,ab,kw.	103
4	(((Fetal or foetal) adj5 fibronectin$) or fFN).ti,ab,kw.	787
5	1 or 2 or 3 or 4	916

Notes: N/A

File name: PartoSure MEDLINE 916 RIS.txt

Database: EMBASE

Host: OVID

Data Parameters: 1974 to 2017 July 21

Date searched: Monday July 24th 2017

Searcher: CC

Search Checked by: JVC

Hits: 1270

#	Searches	Results

1 (PartoSure or Parto Sure or PartoSureR or Parto SureR).ti,ab,kw. 13

2 ((Placental alpha adj5 test$) or PAMG-1).ti,ab,kw. 75

3 (Actim Partus or Actim PartusR or "insulin-like growth factor-binding protein 1 test" or phIGFBP-1 or (IGFBP-1 adj5 test$)).ti,ab,kw. or *actim partus test/ 158

4 (((Fetal or foetal) adj5 fibronectin$) or fFN).ti,ab,kw. 1080

5 1 or 2 or 3 or 4 1270

Notes: N/A

File name: PartoSure Embase 1270 RIS

Database: The Cochrane Library (CDSR, CENTRAL, DARE, HTA & NHS EED)

Host: Wiley Interface

Data Parameters: (CDSR: Issue 7 of 12, July 2017. DARE: Issue 2 of 4, April 2015. CENTRAL: Issue 6 of 12, June 2017. HTA Issue 4 of 4, October 2016; and NHS EEDs Issue 2 of 4, April 2015)

Date searched: Monday July 24th 2017

Searcher: CC

Search Checked by: JVC

Hits: 159* (CDSR 36; DARE 13; CENTRAL 91; HTA 14; NHS EEDs 5)

ID Search Hits

#1 (PartoSure or Parto Sure or PartoSureR or Parto SureR) 17

#2 ((Placental alpha near/5 test*) or (PAMG-1)) 15

#3 (Actim Partus or Actim PartusR or "insulin-like growth factor-binding protein 1 test" or phIGFBP-1 or (IGFBP-1 near/5 test*)) 13

#4 (((Fetal or foetal) near/5 fibronectin*) or fFN) 125

#5 #1 or #2 or #3 or #4 164

Notes: *164 hits were identified. 159 study records were downloaded and five records from the methods register (3) and Cochrane groups register (2) were not downloaded (totaling 164). NHS EED and DARE were searched as an archive since they have not been updated since 2015.

Database: BIOSIS

Host: Clarivate Analytics

Data Parameters: 1969-2017

Date searched: Monday July 24th 2017

Searcher: CC

Search Checked by: JVC

Hits: 806

(PartoSure or Parto Sure or PartoSureR or Parto SureR) n=2

((Placental alpha near/6 test*) or (PAMG-1)) n=25

(Actim Partus or Actim PartusR or "insulin-like growth factor-binding protein 1 test" or phIGFBP-1 or (IGFBP-1 near/6 test*)) n=80

(((Fetal or foetal) near/6 fibronectin*) or fFN) n=716

Notes: N/A

File name: PartoSure BIOSIS 806 RIS

Database: Web of Science

Host: Clarivate Analytics

Data Parameters: 1900-2017

Date searched: Monday July 24th 2017

Searcher: CC

Search Checked by: JVC

Hits: 1358

(PartoSure or Parto Sure or PartoSureR or Parto SureR) n=3

((Placental alpha near/6 test*) or (PAMG-1)) n=45

(Actim Partus or Actim PartusR or "insulin-like growth factor-binding protein 1 test" or phIGFBP-1 or (IGFBP-1 near/6 test*)) n=124

(((Fetal or foetal) near/6 fibronectin*) or fFN) n=1226

Notes: N/A

File name: PartoSure WoS 1358 RIS

Database: CINAHL

Host: EBSCOhost

Data Parameters: 1937-2017

Date searched: Monday July 24th 2017

Searcher: CC

Search Checked by: JVC

Hits: 258

(PartoSure or Parto Sure or PartoSureR or Parto SureR) n=4

((Placental alpha N6 test*) or (PAMG-1)) n=22

(Actim Partus or Actim PartusR or "insulin-like growth factor-binding protein 1 test" or phIGFBP-1 or (IGFBP-1 N6 test*)) n=25

(((Fetal or foetal) N6 fibronectin*) or fFN) n=221

Notes: N/A

File name: PartoSure CINAHL 258 RIS

A1.2 Trial registry searching

Date searched: 29th August 2017

Searcher: SD

Clinical Trials.Gov https://clinicaltrials.gov/ct2/home

Search terms, enumerated as 4 different searches:

Parto Sure

Actim Partus

Fetal fibronectin

Foetal fibronectin

ISRCTN: https://www.isrctn.com/editAdvancedSearch

Parto Sure

Actim Partus

Fetal fibronectin

Foetal fibronectin

A1.3 Web searching

Date Searched: Thursday September 17th 2017

Searcher: CC

The first 50 pages were searched in each instance.

Google

1, PartoSure

2. PartoSure filetype:pdf

3. "Actim Partus"

4. "Actim Partus" filetype:pdf

5. "Fetal fibronectin"

6. "Fetal fibronectin" filetype:pdf

A1.4 Utilities Database Searches

Database(s): Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present

Date searched: Monday September 11th 2017
Search Strategy:

#	Searches	Results
1	exp Obstetric Labor, Premature/	23638
2	((Pre term or preterm or premature or early or immature) adj5 (labo?r or birth$ or childbirth$ or deliver$ or partu$ or baby or babies or child$ or infant$ or toddler$ or postnatal or neonatal)).ti,ab,ot,hw.	188755
3	(PROM or PPROM or PROM or PTB).ti,ab,ot.	7439

4	((Short$ or reduced or multiple) adj4 gestation$).ti,ab,ot.	4711
5	(low$ adj3 birth weight).ti,ab,kw.	26742
6	1 or 2 or 3 or 4 or 5	210020
7	(euroqol or euro qol or eq5d or eq 5d or EQ-5D or EQ-5D-Y or EQ-5D-5L).ti,ab,kw.	7897
8	(sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).ti,ab,kw.	1831
9	(sf10 or sf 10 or short form 10 or shortform 10 or sf ten or sften or shortform ten or short form ten).ti,ab,kw.	102
10	(sf12 or sf 12 or short form 12 or shortform 12 or sf twelve of sftwelve or shortform twelve or short form twelve).ti,ab,kw.	4702
11	(sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).ti,ab,kw.	28
12	(sf20 or sf 20 or short form 20 or shortform 20 or sf twenty of sftwenty or shortform twenty of short form twenty).ti,ab,kw.	390
13	(sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirstysix or shortform thirty six or short form thirty six or short form thirtysix or short form thirty six).ti,ab,kw.	22677
14	(health utilities index$ or (hui or hui1 or hui2 or hui3 or hui4 or hui-4 or hui-1 or hui-2 or hui-3)).ti,ab,kw.	1483
15	("time trade off" or "time tradeoff" or TTO).ti,ab,kw.	1668
16	standard gamble$.ti,ab,kw.	849
17	(QWB or "quality of wellbeing" or "quality of well being" or "quality of well-being" or (index adj3 wellbeing)).ti,ab,kw.	570
18	"discrete choice".ti,ab,kw.	1359
19	(AQoL or "Assessment of Quality of Life").ti,ab,kw.	1679
20	(HYE or HYES or health$1 year$1 equivalent$1).ti,ab,kw.	79
21	((quality adj2 life) or HRQoL or HRQL or QoL or (quality adjusted or adjusted life year$) or QALY* or qald$ or QTIME$ or qale$ or qtime$ or daly*).ti,ab,kw. or Quality of life/ or Quality adjusted life years/	287614

22	(health state or health status).ti,ab,kw. or Health status/ or Health status indicators/	127992
23	Value of Life/	5752
24	((utilit$ or disutilit$) adj3 (health$ or score$1 or scoring or valu$ or measur$ or evaluat$ or scale$1 or instrument$1 or weight or weights or weighting or information or data or unit or units or life or estimat$ or elicit$ or disease$ or mean or cost$ or expenditure$1 or gain or gains or loss or losses or lost or analysis or index$ or indices or overall or reported or calculat$ or range$ or increment$ or state or states or status)).ti,ab,kw.	29499
25	7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24	422994
26	(Parental Stressor Scale or PSS:NICU or Edinburgh Postpartum Depression Scale or EPDS or Spielberger State-Trait Anxiety Inventory or STAI or "Family Adaptability and Cohesion Evaluation Scale II" or FACES II or "The Impact of Bronchiolitis Hospitalization Questionnaire" or IBHQ or "Preschool Children Quality of Life Questionnaire" or TNO AZL or TAPQOL or "Pediatric Quality of Life Inventory" or "Child Health Questionnaire" or CHQ or "The Preterm Birth Experience and Satisfaction Scale").ti,ab,kw.	6377
27	25 or 26	427211
28	6 and 27	4129
29	limit 28 to english language	3700

Database(s): Embase 1974 to 2017 September 08
Search Strategy:

#	Searches	Results
1	exp Obstetric Labor, Premature/	40189
2	((Pre term or preterm or premature or early or immature) adj5 (labo?r or birth$ or childbirth$ or deliver$ or partu$ or baby or babies or child$ or infant$ or toddler$ or postnatal or neonatal)).ti,ab,ot,hw.	208972
3	(PROM or PPROM or PROM or PTB).ti,ab,ot.	10995
4	((Short$ or reduced or multiple) adj4 gestation$).ti,ab,ot.	6592
5	(low$ adj3 birth weight).ti,ab,kw.	33872

6	1 or 2 or 3 or 4 or 5	238680
7	(euroqol or euro qol or eq5d or eq 5d or EQ-5D or EQ-5D-Y or EQ-5D-5L).ti,ab,kw.	13620
8	(sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).ti,ab,kw.	1929
9	(sf10 or sf 10 or short form 10 or shortform 10 or sf ten or sften or shortform ten or short form ten).ti,ab,kw.	149
10	(sf12 or sf 12 or short form 12 or shortform 12 or sf twelve of sftwelve or shortform twelve or short form twelve).ti,ab,kw.	7188
11	(sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).ti,ab,kw.	47
12	(sf20 or sf 20 or short form 20 or shortform 20 or sf twenty of sftwenty or shortform twenty of short form twenty).ti,ab,kw.	381
13	(sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirstysix or shortform thirty six or short form thirty six or short form thirtysix or short form thirty six).ti,ab,kw.	33894
14	(health utilities index$ or (hui or hui1 or hui2 or hui3 or hui4 or hui-4 or hui-1 or hui-2 or hui-3)).ti,ab,kw.	2047
15	("time trade off" or "time tradeoff" or TTO).ti,ab,kw.	2242
16	standard gamble$.ti,ab,kw.	992
17	(QWB or "quality of wellbeing" or "quality of well being" or "quality of well-being" or (index adj3 wellbeing)).ti,ab,kw.	680
18	"discrete choice".ti,ab,kw.	1938
19	(AQoL or "Assessment of Quality of Life").ti,ab,kw.	2465
20	(HYE or HYES or health$1 year$1 equivalent$1).ti,ab,kw.	135
21	((quality adj2 life) or HRQoL or HRQL or QoL or (quality adjusted or adjusted life year$) or QALY* or qald$ or QTIME$ or qale$ or qtime$ or daly*).ti,ab,kw. or Quality of life/ or Quality adjusted life years/	464722
22	(health state or health status).ti,ab,kw. or Health status/ or Health status indicators/	136100
23	Value of Life/	124596

24	((utilit$ or disutilit$) adj3 (health$ or score$1 or scoring or valu$ or measur$ or evaluat$ or scale$1 or instrument$1 or weight or weights or weighting or information or data or unit or units or life or estimat$ or elicit$ or disease$ or mean or cost$ or expenditure$1 or gain or gains or loss or losses or lost or analysis or index$ or indices or overall or reported or calculat$ or range$ or increment$ or state or states or status)).ti,ab,kw.	42607
25	7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24	726333
26	(Parental Stressor Scale or PSS:NICU or Edinburgh Postpartum Depression Scale or EPDS or Spielberger State-Trait Anxiety Inventory or STAI or "Family Adaptability and Cohesion Evaluation Scale II" or FACES II or "The Impact of Bronchiolitis Hospitalization Questionnaire" or IBHQ or "Preschool Children Quality of Life Questionnaire" or TNO AZL or TAPQOL or "Pediatric Quality of Life Inventory" or "Child Health Questionnaire" or CHQ or "The Preterm Birth Experience and Satisfaction Scale").ti,ab,kw.	9565
27	25 or 26	732723
28	6 and 27	8472
29	limit 28 to english language	

NHS EEDs Issue 2 of 4, April 2015

Search Name:

Date Run: 11/09/17 12:54:41.877

Description:

ID	Search	Hits
#1	MeSH descriptor: [Obstetric Labor, Premature] explode all trees	1317
#2	((Pre term or preterm or premature or early or immature) near/3 (labor or labour or birth* or childbirth* or deliver* or partu* or baby or babies or child* or infant* or toddler* or postnatal or neonatal))	15146
#3	(PROM or PPROM or PROM or PTB)	775
#4	((Short* or reduced or multiple) near/4 gestation*) ((Short* or reduced or multiple) near/4 gestation*)	563

#5 (low* near/3 birth weight) (low* near/3 birth weight) 4485

#6 #1 or #2 or #3 or #4 or #5 17627

NHS EEDs n=250

ScHARR HUD was handsearched

A1.5 Adverse events Database searches

Medline

1. exp Obstetric Labor, Premature/

2. ((Pre term or preterm or premature or early or immature) adj5 (labo?r or birth$ or childbirth$ or deliver$ or partu$ or baby or babies or child$ or infant$ or toddler$ or postnatal or neonatal)).ti,ab,ot,hw.

3. (PROM or PPROM or PROM or PTB).ti,ab,ot.

4. ((Short$ or reduced or multiple) adj4 gestation$).ti,ab,ot.

5. (low$ adj3 birth weight).ti,ab,kw.

6. 1 or 2 or 3 or 4 or 5

7. (cost$ or healthcare utilisation or healthcare utilization or expend$ or price$ or pricing or budget$ or value$).ti,ab,kw.

8. ((neonat$ or newborn$) and (mortality or death)).ti,ab,kw.

9. respiratory distress syndrome.ti,ab,kw.

10. intraventricular haemorrhage.ti,ab,kw.

11. 8 or 9 or 10

12. 6 and 7 and 11

13. limit 15 to english language

Appendix 2. Supplementary tables

Table 49 Included systematic reviews

1.	The Actim™ Partus versus the TLIIQ® System as rapid response tests to aid in diagnosing preterm labour in symptomatic women (Structured abstract). Health Technology Assessment Database [Internet]. 2008; (4). Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32008100231/frame.html.
2.	Post policy implementation review (PPIR) of rapid fetal fibronectin testing for preterm labour in Alberta (Structured abstract). Health Technology Assessment Database [Internet]. 2015; (4). Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32015001020/frame.html.
3.	Conde-Agudelo A, Romero R. Cervical phosphorylated insulin-like growth factor binding protein-1 test for the prediction of preterm birth: a systematic review and metaanalysis. American Journal of Obstetrics & Gynecology. 2016;214(1):57-73.
4.	Honest H, Bachmann LM, Gupta JK, Kleijnen J, Khan KS. Accuracy of cervicovaginal fetal fibronectin test in predicting risk of spontaneous preterm birth: systematic review. BMJ: British Medical Journal (International Edition). 2002;325(7359):301-4.
5.	Leitich H, Kaider A. Fetal fibronectin - how useful is it in the prediction of preterm birth? Bjog-an International Journal of Obstetrics and Gynaecology. 2003;110:66-70.
6.	Lucaroni F, Morciano L, Rizzo G, F DA, Buonuomo E, Palombi L, et al. Biomarkers for predicting spontaneous preterm birth: an umbrella systematic review. Journal of Maternal-Fetal and Neonatal Medicine. 2017:1-9.
7.	Menon R, Torloni MR, Voltolini C, Torricelli M, Merialdi M, Betran AP, et al. Biomarkers of Spontaneous Preterm Birth: An Overview of The Literature in the Last Four Decades. Reproductive Sciences. 2011;18(11):1046-70.
8.	Sanchez-Ramos L, Delke I, Kaunitz A. Cervico-vaginal fetal fibronectin as a short-term predictor of preterm birth in symptomatic patients: A meta-analysis of diagnostic accuracy. American journal of obstetrics and gynecology. 2007;197(6):S198-S.
9.	Sanchez-Ramos L, Delke I, Zamora J, Kaunitz AM. Fetal fibronectin as a short-term predictor of preterm birth in symptomatic patients: a meta-analysis. Obstetrics & Gynecology. 2009;114(3):631-40.
10.	Smith V, Devane D, Begley CM, Clarke M, Higgins S. A systematic review and quality assessment of systematic reviews of fetal fibronectin and transvaginal length for predicting preterm birth. European Journal of Obstetrics Gynecology and Reproductive Biology. 2007;133(2):134-42.
11.	Vis JY, Wilms FF, Oudijk MA, Bossuyt PM, van der Post JA, Grobman WA, et al. Why were the results of randomized trials on the clinical utility of fetal fibronectin negative? A systematic review of their study designs. American Journal of Perinatology. 2011;28(2):145-50.

Table 50 Included citations

1	Abo El-Ezz AE, Askar AE. Predictive value of phosphorylated insulin-like growth factor binding protein-1 (PIGFBP-1) (bedside test) in preterm labor. Journal of the Egyptian Society of Parasitology. 2014;44(2):525-30.	FT
2	Altinkaya O, Gungor T, Ozat M, Danisman N, Mollamahmutoglu L. Cervical phosphorylated insulin-like growth factor binding protein-1 in prediction of preterm delivery. Archives of Gynecology and Obstetrics. 2009;279(3):279-83.	FT
3	Azlin MI, Bang HK, An LJ, Mohamad SN, Mansor NA, Yee BS, et al. Role of phIGFBP-1 and ultrasound cervical length in predicting pre-term labour. Journal of Obstetrics & Gynaecology. 2010;30(5):456-9.	FT
4	Azlin MN, Kee BH, Low JA, Mohamad SN, Mansor NA, Bee SY, et al. Role of phigfbp-1 and cervical length in predicting preterm labor. Journal of Maternal-Fetal and Neonatal Medicine. 2010;23:100.	Ab
5	Bolotskikh V, Borisova V. Combined value of placental alpha microglobulin-1 detection and cervical length via transvaginal ultrasound in the diagnosis of preterm labor in symptomatic patients. Journal of Obstetrics and Gynaecology Research. 2017.	FT
6	Brik M, Hernandez AIM, Pedraz CC, Perales A. Phosphorylated insulin-like growth factor binding protein-1 and cervical measurement in women with threatening preterm birth. Acta Obstetricia Et Gynecologica Scandinavica. 2010;89(2):268-74.	FT
7	Bruijn M, Kamphuis E, Hoesli I, de Tejada BM, Loccufier A, Jacquemyn Y, et al. Does quantitative fetal fibronectin testing add to cervical length measurement and qualitative fetal fibronectin testing; (I) Identification of low risk women with threatened preterm labor (EUFIS study). American Journal of Obstetrics and Gynecology. 2015;212(1, Suppl. S):S190-S1.	Ab
8	Bruijn M, Kamphuis E, Hoesli I, de Tejada BM, Loccufier A, Jacquemyn Y, et al. Does quantitative fetal fibronectin testing add to cervical length measurement and qualitative fetal fibronectin testing; (II) absolute probability of preterm delivery < 7 days in women with threatened preterm labor (EUFIS study). American Journal of Obstetrics and Gynecology. 2015;212(1):S191-S2.	Ab
9	Bruijn M, van Baaren G-J, Vis J, van Straalen J, Wilms F, Oudijk M, et al. Does quantitative fetal fibronectin testing improve the prediction of spontaneous preterm delivery as compared to qualitative fetal fibronectin testing in symptomatic women: a post-hoc analysis. American Journal of Obstetrics and Gynecology. 2014;210(1, Suppl. S):S364.	Ab
10	Bruijn M, van Baaren G-J, Vis J, van Straalen J, Wilms F, Oudijk M, et al. Comparison of the Actim Partus test and fetal fibronectin test in combination with cervical length in the prediction of spontaneous preterm delivery in symptomatic women: a post-hoc analysis. American Journal of Obstetrics and Gynecology. 2014;210(1, Suppl. S):S363-S4.	Ab
11	Bruijn M, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM, et al. Quantitative fetal fibronectin testing in combination with cervical length measurement in the prediction of spontaneous preterm delivery in symptomatic women. BJOG: An International Journal of Obstetrics & Gynaecology. 2016;123(12):1965-71.	FT
12	Bruijn MM, Kamphuis EI, Hoesli IM, Martinez de Tejada B, Loccufier AR, Kuhnert M, et al. The predictive value of quantitative fibronectin testing in combination with cervical length measurement in symptomatic women. American Journal of Obstetrics & Gynecology. 2016;215(6):793.e1-.e8.	FT
13	Bruijn MM, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM, et al. Comparison of the Actim Partus test and the fetal fibronectin test in the prediction of spontaneous preterm birth in symptomatic women undergoing cervical length measurement. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2016;206:220-4.	FT
14	Cooper S, Lange I, Wood S, Tang S, Miller L, Ross S. Diagnostic accuracy of rapid phIGFBP-I assay for predicting preterm labor in symptomatic patients. Journal of Perinatology. 2012;32(6):460-5.	FT
15	Danti L, Prefumo F, Lojacono A, Corini S, Testori A, Frusca T. The combination of short cervical length and phIGFBP-1 in the prediction of preterm delivery in symptomatic women. Journal of maternal-fetal & neonatal medicine [Internet]. 2011; 24(10):[1262-6 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/086/CN-00890086/frame.html.	FT
16	Eroglu D, Yanik F, Oktem M, Zeyneloglu HB, Kuscu E. Prediction of preterm delivery among women with threatened preterm labor. Gynecologic and Obstetric Investigation. 2007;64(2):109-16.	FT

17	Goyal M, Kriplani A, Kachhawa G, Badiger S. Prediction of preterm labor by a rapid bedside test detecting phosphorylated insulin-like growth factor-binding protein 1 in cervical secretions. International Journal of Gynecology & Obstetrics. 2016;134(2):165-8.	FT
18	Hadzi Lega M, Daneva A, Girevski V. EP18.02: Comparison of partosure (PAMG-1) and actim partus (phIGFBP-1) for the prediction of preterm delivery in patients with preterm labour and a short cervix. Ultrasound in Obstetrics & Gynecology. 2016;48:345-.	Ab
19	Hadzi Lega M, Hellmeyer L, Hanns H, Josefine M, Poposka A, Daneva Markova A. Comparison of partosure (PAMG-1) and actim partus (phlgfbp-1) for the prediction of preterm delivery in patients with preterm labor and a short cervix. Journal of Maternal-Fetal and Neonatal Medicine. 2016;29:55.	Ab
20	Hadzi-Lega M, Maier JT, Helmer H, Hellmeyer L, Markova AD, Poposka A. Comparison of PAMG-1 and phIGFBP-1 Tests for the Prediction of Preterm Delivery in Patients with Preterm Labor. Open Journal of Obstetrics and Gynecology. 2017;Vol.07No.03:11.	FT
21	Lembet A, Eroglu D, Ergin T, Kuscu E, Haberal A, Gaddipati S. A new bed-side test for the prediction of preterm delivery: Phosphorylated insulin-like growth factor binding protein-1 in cervical secretions. American Journal of Obstetrics and Gynecology. 2001;185(6 Supplement):S147.	Ab
22	Lembet A, Eroglu D, Ergin T, Kuscu E, Zeyneloglu H, Batioglu S, et al. New rapid bed-side test to predict preterm delivery: phosphorylated insulin-like growth factor binding protein-1 in cervical secretions. Acta Obstetricia Et Gynecologica Scandinavica. 2002;81(8):706-12.	FT
23	Nikolova T, Bayev O, Nikolova N, Di Renzo GC. Evaluation of a novel pamg-1 test (partosurea TTD test) to predict time to delivery in patients with preterm labor. Journal of Perinatal Medicine Conference: 11th World Congress of Perinatal Medicine. 2013;41(no pagination).	Ab
24	Nikolova T, Bayev O, Nikolova N, Di Renzo GC. Evaluation of a novel placental alpha microglobulin-1 (PAMG-1) test to predict spontaneous preterm delivery. Journal of Perinatal Medicine. 2014;42(4):473-7.	FT
25	Nikolova T, Bayev O, Nikolova N, Di Renzo GC. Comparison of a novel test for placental alpha microglobulin-1 with fetal fibronectin and cervical length measurement for the prediction of imminent spontaneous preterm delivery in patients with threatened preterm labor. Journal of Perinatal Medicine. 2015;43(4):395-402.	FT
26	Riboni F, Vitulo A, Dell'avanzo M, Plebani M, Battagliarin G, Paternoster D. Biochemical markers predicting pre-term delivery in symptomatic patients: phosphorylated insulin-like growth factor binding protein-1 and fetal fibronectin. Archives of Gynecology and Obstetrics. 2011;284(6):1325-9.	FT
27	Tanir HM, Sener T, Yildiz Z. Cervical phosphorylated insulin-like growth factor binding protein-1 for the prediction of preterm delivery in symptomatic cases with intact membranes. Journal of Obstetrics and Gynaecology Research. 2009;35(1):66-72.	FT
28	Ting H-S, Chin P-S, Yeo GS, Kwek K. Comparison of bedside test kits for prediction of preterm delivery: Phosphorylated insulin-like growth factor binding protein-1 (pIGFBP-1) test and fetal fibronectin test. Annals Academy of Medicine Singapore. 2007;36(6):399-402.	FT
29	Tripathi R, Tyagi S, Mala YM, Singh N, Pandey NB, Yadav P. Comparison of rapid bedside tests for phosphorylated insulin-like growth factor-binding protein 1 and fetal fibronectin to predict preterm birth. International Journal of Gynecology & Obstetrics. 2016;135(1):47-50.	FT
30	Vishwekar PS, Chauhan AR, Turakhia N. Prediction of preterm delivery with a novel bedside test. 2017. 2017;6(8):6.	FT
31	Werlen S, Raia T, Di Bartolomeo A, Chauleur C. [Preterm labor: Reproducibility of detection test of PAMG-1 before and after digital examination, and transvaginal ultrasound cervical length]. Gynecologie, obstetrique & fertilite. 2015;43(10):640-5.	FT

Key: Ab, abstract; FT, full text;

Table 51 Alere (Actim Partus) Submitted citations

No.	Citation	Reason for Exclusion	Further detail
1.	Adeyemi O, Osoba L. The role of phosphorylated insulin-like growth factor binding protein-1 in predicting pre-term labour in twin pregnancies. J Obstet Gynaecol. 2010;30(6):571-3.	Population	Twin pregnancies
2.	Akercan F, Kazandi M, Sendag F, Cirpan T, Mgoyi L, Terek MC, et al. Value of cervical phosphorylated insulinlike growth factor binding protein-1 in the prediction of preterm labor. J Reprod Med. 2004;49(5):368-72.	Comparator	Outcome at 37 weeks
3.	Altinkaya O, Gungor T, Ozat M, Danisman N, Mollamahmutoglu L. Cervical phosphorylated insulin-like growth factor binding protein-1 in prediction of preterm delivery. Arch Gynecol Obstet. 2009;279(3):279-83.	Included	
4.	Azlin MI, Bang HK, An LJ, Mohamad SN, Mansor NA, Yee BS, et al. Role of phIGFBP-1 and ultrasound cervical length in predicting pre-term labour. J Obstet Gynaecol. 2010;30(5):456-9.	Included	
5.	Balic D, Latifagic A, Hudic I. Insulin-like growth factor-binding protein-1 (IGFBP-1) in cervical secretions as a predictor of preterm delivery. J Matern Fetal Neonatal Med. 2008;21(5):297-300.	Population	Asymptomatic population
6.	Bittar RE, Da Fonseca EB, De Carvalho MHB, Martinelli S, Zugaib M. Predicting preterm delivery in asymptomatic patients with prior preterm delivery by measurement of cervical length and phosphorylated insulin-like growth factor-binding protein-1. Ultrasound Obst Gyn. 2007;29(5):562-7.	Population	Asymptomatic population
7.	Brik M, Hernandez AIM, Pedraz CC, Perales A. Phosphorylated insulin-like growth factor binding protein-1 and cervical measurement in women with threatening preterm birth. Acta Obstet Gyn Scan. 2010;89(2):268-74.	Included	
8.	Bruijn MMC, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM, et al. Comparison of the Actim Partus test and the fetal fibronectin test in the prediction of spontaneous preterm birth in symptomatic women undergoing cervical length measurement. Eur J Obstet Gyn R B. 2016;206:220-4.	Included	
9.	Cooper S, Lange I, Wood S, Tang S, Miller L, Ross S. Diagnostic accuracy of rapid phIGFBP-I assay for predicting preterm labor in symptomatic patients. Journal of perinatology: official journal of the California Perinatal Association. 2011;32(6):460-5.	Included	
10.	Danti L, Prefumo F, Lojacono A, Corini S, Testori A, Frusca T. The combination of short cervical length and phIGFBP-1 in the prediction of preterm delivery in symptomatic women. Journal of Maternal-Fetal & Neonatal Medicine. 2011;24(10):1262-6.	Included	
11.	Dogl M, Skogvoll E, Heimstad R. Cervical insulin-like growth factor binding protein-1 (IGFBP-1) to predict spontaneous onset of labor and induction to delivery interval in post-term pregnancy. Acta Obstet Gyn Scan. 2011;90(1):57-62.	Population	Post-term pregnancies
12.	Elizur SE, Yinon Y, Epstein GS, Seidman DS, Schiff E, Sivan E. Insulin-like growth factor binding protein-1 detection in preterm labor: Evaluation of a bedside test. Am J Perinat. 2005;22(6):305-9.	Population	Not clear if multi-foetal also outcome at 35 weeks
13.	Eroglu D, Yanik F, Oktem M, Zeyneloglu HB, Kuscu E. Prediction of preterm delivery among women with threatened preterm labor. Gynecol Obstet Inves. 2007;64(2):109-16.	Included	
14.	Hadzi-Lega M, Markova AD, Stefanovic M, Tanturovski M. Correlation of cervical length, fetal fibronectin, phIGFBP-1, and cytokines in spontaneous preterm birth up to 14 days from sampling. J Perinat Med. 2014;43(5):545-51.	Comparator	Outcome at 14 days
15.	Kekki M, Kurki T, Karkkainen T, Hiilesmaa V, Paavonen J, Rutanen EM. Insulin-like growth factor-binding protein-1 in cervical secretion as a predictor of preterm delivery. Acta Obstet Gyn Scan. 2001;80(6):546-51.	Outcome	Unclear what time reference is, also includes multifetal
16.	Kekki M, Kurki T, Paavonen J, Rutanen EM. Insulin-like growth factor binding protein-1 in cervix as a marker of infectious complications in pregnant women with bacterial vaginosis. Lancet. 1999;353(9163):1494-.	Study Design	Letter

No.	Citation	Reason for Exclusion	Further detail
17.	Khambay H, Bolt LA, Chandiramani M, De Greeff A, Filmer JE, Shennan AH. The Actim Partus test to predict pre-term birth in asymptomatic high-risk women. J Obstet Gynaecol. 2012;32(2):132-4.	Population	Asymptomatic
18.	Kosinska-Kaczynska K, Bomba-Opon D, Bobrowska K, Kozlowski S, Brawura-Biskupski-Samaha R, Szymusik I, et al. Phosphorylated IGFBP-1 in predicting successful vaginal delivery in post-term pregnancy. Arch Gynecol Obstet. 2015;292(1):45-52.	Population	Post-term population
19.	Kwek K, Khi C, Ting HS, Yeo GSH. Evaluation of a bedside test for phosphorylated insulin-like growth factor binding protein-1 in preterm labour. Ann Acad Med Singap. 2004;33(6):780-3.	Population	Unclear if includes multifetal
20.	Latifagic A, Balic D, Fatusic Z, Hudic I, Kapidzic M, Habibovicd A. Insulin-like growth factor-binding protein-1 (IGFBP-1) in cervical secretions in women with symptoms of preterm delivery. Med Glas. 2008;5(2):121-4.	Outcome	Unclear what time reference is
21.	Lembet A, Eroglu D, Ergin T, Kuscu E, Zeyneloglu H, Batioglu S, et al. New rapid bed-side test to predict preterm delivery: phosphorylated insulin-like growth factor binding protein-1 in cervical secretions. Acta Obstet Gynecol Scand. 2002;81(8):706-12.	Included	
22.	Mesic Ethogic L, Micic D, Omeragic F, Kovac R, Fazlagic S. IGFBP-1 marker of cervical ripening and predictor of preterm birth. Med Glas (Zenica). 2016;13(2):118-24.	Population	Asymptomatic population
23.	Nuutila M, Hiilesmaa V, Karkkainen T, Ylikorkala O, Rutanen EM. Phosphorylated isoforms of insulin-like growth factor binding protein-1 in the cervix as a predictor of cervical ripeness. Obstet Gynecol. 1999;94(2):243-9.	Population	Term population
24.	Park O-R et al. Usefulness of phosphorylated insulin-like growth factor binding protein-1 for prediction of preterm delivery. Korean J Obstet Gynecol (2003) 46:1378-1383. Korean J Obstet Gynecol. 2003;46:1378-83.	Language	
25.	Paternoster DM, Muresan D, Vitulo A, Serena A, Battagliarin G, Dell'Avanzo M, et al. Cervical phIGFBP-1 in the evaluation of the risk of preterm delivery. Acta Obstet Gyn Scan. 2007;86(2):151-5.	Outcome	Outcome at 34 weeks
26.	Rahkonen L. Prediction of pre-term delivery with phosphorylated insulin-like growth factor-binding protein-1. European Obstetrics & Gynecology. 2011;6:3-7.	Study Design	Review
27.	Riboni F, Vitulo A, Dell'avanzo M, Plebani M, Battagliarin G, Paternoster D. Biochemical markers predicting pre-term delivery in symptomatic patients: phosphorylated insulin-like growth factor binding protein-1 and fetal fibronectin. Arch Gynecol Obstet. 2011;284(6):1325-9.	Included	
28.	Riboni F, Vitulo A, Plebani M, Dell'avanzo M, Battagliarin G, Paternoster D. Combination of biochemical markers in predicting pre-term delivery. Arch Gynecol Obstet. 2012;285(1):61-6.	Population	Asymptomatic
29.	Rolnik DL, Bittar RE, de Carvalho MH, Zugaib M, Francisco RP. [Preterm birth prediction: sequential evaluation of the cervix and the test for phosphorylated protein-1 linked to insulin-like growth factor]. Revista brasileira de ginecologia e obstetricia : revista da Federacao Brasileira das Sociedades de Ginecologia e Obstetricia. 2013;35(9):394-400.	Population	Asymptoamtic
30.	Shine BK, et al. Insulin-like growth factor-binding protein-1 in cervical secretion as a predictor of preterm delivery. Korean J Obstet Gynecol. 2001;44:2250-6.	Language	
31.	Tanir HM, Sener T, Yildiz Z. Cervical phosphorylated insulin-like growth factor binding protein-1 for the prediction of preterm delivery in symptomatic cases with intact membranes. J Obstet Gynaecol Res. 2009;35(1):66-72.	Included	
32.	Ting HS, Chin PS, Yeo GS, Kwek K. Comparison of bedside test kits for prediction of preterm delivery: Phosphorylated insulin-like growth factor binding protein-1 (pIGFBP-1) test and fetal fibronectin test. Ann Acad Med Singap. 2007;36(6):399-402.	Included	
33.	Tripathi R, Tyagi S, Mala YM, Singh N, Pandey NB, Yadav P. Comparison of rapid bedside tests for phosphorylated insulin-like growth factor-binding protein 1 and fetal fibronectin to predict preterm birth. Int J Gynaecol Obstet. 2016;135(1):47-50.	Included	

No.	Citation	Reason for Exclusion	Further detail
34.	Vallikkannu N, Lam WK, Omar SZ, Tan PC. Insulin-like growth factor binding protein 1, Bishop score, and sonographic cervical length: tolerability and prediction of vaginal birth and vaginal birth within 24 hours following labour induction in nulliparous women. BJOG. 2017;124(8):1274-83.	Population	Term women, labour induction

Table 52 Hologic (fFN) Submitted citations

No.	Citation	Reason for Exclusion	Further detail
1	Abbott D, Radford S, Foster C, Vousden N, Shennan A. Longitudinal trend of quantitative fetal fibronectin in the prediction of delivery following insertion of a rescue cerclage. J Obstet Gynaecol. 2013 May;33(4):414-5.	Study Design	Case Study
2	Abbott DS, Hezelgrave NL, Seed PT, et al. Quantitative fetal fibronectin to predict preterm birth in asymptomatic women at high risk. AH.Obstet Gynecol. 2015 May;125(5):1168-76.	Population	Asymptomatic population
3	Abbott DS, Radford SK, Seed PT, Tribe RM, Shennan AH. Evaluation of a quantitative fetal fibronectin test for spontaneous preterm birth in symptomatic women. Am J Obstet Gynecol. 2013 Feb;208(2):122.e1-6.	Comparator	Outcome at 14 days
4	Anderson-Knight HE, Hezelgrave NL, Shennan AH. J Obstet Gynaecol. Spontaneous resolution of a midtrimester dilated cervix with expectant management guided by quantitative foetal fibronectin results. J Obstet Gynaecol. 2015;35(7):766-7.	Study Design	Case Study
5	Bolt LA, Chandiramani M, De Greeff A, Seed P, Shennan AH. Does fetal fibronectin testing change patient management in women at risk of preterm labour? Eur J Obstet Gynecol Reprod Biol. 2009 Oct;146(2):180-3.	Comparator	Outcome at 14 days
6	Bolt LA, Chandiramani M, De Greeff A, Seed PT, Kurtzman J, Shennan AH. The value of combined cervical length measurement and fetal fibronectin testing to predict spontaneous preterm birth in asymptomatic high-risk women. J Matern Fetal Neonatal Med. 2011 Jul;24(7):928-32.	Population	Asymptomatic population
7	Bolt LA, Morrison K, Shennan AH. The use of fetal fibronectin testing and cervical length measurement in the prediction of delivery of triplet pregnancies. Eur J Obstet Gynecol Reprod Biol. 2012 Oct;164(2):236-7.	Population	Triplet pregnancies
8	Bruijn M, Vis J, Wilms F, et al. Quantitative fetal fibronectin testing in combination with cervical length measurement in the prediction of spontaneous preterm delivery in symptomatic women. BJOG: An International Journal of Obstetrics & Gynaecology. 2016;123:1965-1971.	Included	
9	Bruijn MM, Kamphuis EI, Hoesli IM, et al. The predictive value of quantitative fibronectin testing in combination with cervical length measurement in symptomatic women. Am J Obstet Gynecol. 2016 Dec;215(6):793.e1-793.	Included	
10	Centra M, Coata G, Picchiassi E, et al. Evaluation of quantitative fFn test in predicting the risk of preterm birth. J Perinat Med. 2017 Jan 1;45(1):91-98.	Comparator	Outcome at 14 days
11	Fiorini F, Isted A, Hezelgrave NL, Shennan AH. Quantitative fetal fibronectin predicts preterm birth in women with bulging fetal membranes. Eur J Obstet Gynecol Reprod Biol. 2016 Aug;203:127-31.	Comparator	Outcome at 14 days
12	Foster C, Shennan AH. Fetal fibronectin as a biomarker of preterm labor: a review of the literature and advances in its clinical use. Biomark Med. 2014;8(4):471-84.	Study Design	Literature review
13	Gibson S, Hezelgrave NL, Shennan AH. Management of vasa praevia: a potential role for cervical length and quantitative fetal fibronectin measurement. J Obstet Gynaecol. 2013 Nov;33(8):905-6.	Study Design	Case Study
14	Goepfert AR, Goldenberg RL, Mercer B et al. The preterm prediction study: quantitative fetal fibronectin values and the prediction of spontaneous preterm birth: the National Institute of Child Health and Human Development maternal-fetal medicine units network. Am J Obstet Gynecol 2000;183:1480-3.	Population	Asymptomatic population

No.	Citation	Reason for Exclusion	Further detail
15	Goldenberg RL, Iams JD, Mercer BM, et al. The preterm prediction study: the value of new vs standard risk factors in predicting early and all spontaneous preterm births; NICHD MFMU network. Am J Public Health 1998:88:233-8.183:469-75.	Population	Asymptomatic population
16	Goldenberg RL, Klebanoff M, Carey JC et al. Vaginal fetal fibronectin measurements from 8 to 22 weeks' gestation and subsequent spontaneous preterm birth. Am J Obstet Gynecol 2000;	Population	Asymptomatic population
17	Goldenberg RL, Mercer BM, Meis PJ, Copper RL, Das A, McNellis D. The preterm prediction study: fetal fibronectin testing and spontaneous preterm birth; NICHD maternal fetal medicine units network. Obstet Gynecol 1996; 87:643-648.	Population	Asymptomatic population
18	Golic M, Siedentopf JP, Pauly F, Hinkson L, Henrich W, Tucher E. Influence of transvaginal ultrasound examination on quantitative vaginal fibronectin measurements: a prospective evaluation study. J Perinat Med. 2017 Jan 1;45(1):85-89.	Outcome	Does not report test accuracy data
19	Hezelgrave NL, Kuhrt K, Cottam K, Seed PT, Tribe RM, Shennan AH. The effect of blood staining on cervicovaginal quantitative fetal fibronectin concentration and prediction of spontaneous preterm birth. Eur J Obstet Gynecol Reprod Biol. 2017 Jan;208:103-108.	Population	Asymptomatic population
20	Hezelgrave NL, Shennan AH. Quantitative fetal fibronectin to predict spontaneous preterm birth: a review. Womens Health (Lond). 2016 Jan;12(1):121-8.	Population	Asymptomatic population
21	Hezelgrave, NL, et al. Quantitative Fetal Fibronectin at 18 Weeks of Gestation to Predict Preterm Birth in Asymptomatic High-Risk Women. Obstet *Gynecol*, 127(2): 255-63 (2016)	Population	Asymptomatic population
22	Jwala S, Tran TL, Terenna C, et al. Evaluation of additive effect of quantitative fetal fibronectin to cervical length for prediction of spontaneous preterm birth among asymptomatic low-risk women. Acta Obstetricia et Gynecologica Scandinavica. 2016;95:948-955.	Population	Asymptomatic population
23	Kuhrt K, Hezelgrave N, Foster C, Seed PT, Shennan AH. Development and validation of a tool incorporating quantitative fetal fibronectin to predict spontaneous preterm birth in symptomatic women. Ultrasound Obstet Gynecol. 2016 Feb;47(2):210-6	Comparator	Outcome at 14 days
24	Kuhrt K, Smout E, Hezelgrave N, Seed PT, Carter J, Shennan AH. Development and validation of a tool incorporating cervical length and quantitative fetal fibronectin to predict spontaneous preterm birth in asymptomatic high-risk women. Ultrasound Obstet Gynecol. 2016 Jan;47(1):104-9.	Population	Asymptomatic population
25	Kuhrt K, Unwin C, Hezelgrave N, Seed P, Shennan A. Endocervical and high vaginal quantitative fetal fibronectin in predicting preterm birth. J Matern Fetal Neonatal Med. 2014 Oct;27(15):1576-9.	Population	Asymptomatic population
26	Kurtzman JT, Chandiramani M, Briley A et al. Quantitative fetal fibronectin screening in asymptomatic high-risk patients and the spectrum of risk for recurrent preterm delivery. Am J Obstet Gynecol 2009;200:263.e1-263.e6.	Population	Asymptomatic population
27	Lu, GC, et al. Vaginal fetal fibronectin levels and spontaneous preterm birth in symptomatic women. Obstet Gynecol, 97(2): 225-8 (2001).	Study Design	ELISA test
28	McLaren, JS, et al. Prediction of spontaneous preterm birth using quantitative fetal fibronectin after recent sexual intercourse. Am J Obstet Gynecol, 212(1): 89.e1-5 (2015).	Population	Asymptomatic population
29	Min, J, et al. Ability of a preterm surveillance clinic to triage risk of preterm birth: a prospective cohort study. Ultrasound Obstet Gynecol, 48(1): 38-42 (2016).	Population	Asymptomatic population
30	Ridout A, Carter J, Shennan A. Clinical utility of quantitative fetal fibronectin in preterm labour. BJOG. 2016 Nov;123(12):1972.	Study Design	Letter
31	Ross GN, Ridout AE, Shennan AH. Optimal clinical risk prediction can be achieved by combining quantitative fetal fibronectin and cervical length, and avoiding thresholds. Acta Obstet Gynecol Scand. 2016 Aug;95(8):956.	Study Design	Letter
32	Schindhelm RK, Hoogenberg J, de Vos MT, Tegelaers FP. Analytical performance of quantitative fetal fibronectin assay. Ultrasound Obstet Gynecol. 2016 Jan;47(1):127.	Study Design	Letter

No.	Citation	Reason for Exclusion	Further detail
33	Van der Krogt, L, et al. Prediction of spontaneous preterm birth using fetal fibronectin in women with a low-lying placenta. J Matern Fetal Neonatal Med, 30(3): 313-316 (2017).	Population	Asymptomatic population
34	Vandermolen, BI, et al. Quantitative fetal fibronectin and cervical length to predict preterm birth in asymptomatic women with previous cervical surgery. Am J Obstet Gynecol, 215(4): 480.e1-480.e10 (2016).	Population	Asymptomatic population
35	Watson HA, Carter J, Seed PT, Tribe RM, Shennan AH. The QUIPP app: a safe alternative to a treat-all strategy for threatened preterm labour. Ultrasound Obstet Gynecol. 2017 Apr 24.	Abstract	Use of phone application
36	Zhou MX, Zhou J, Bao Y, Chen YQ, Cai C. Evaluation of the ability of cervical length and fetal fibronectin measurement to predict preterm delivery in asymptomatic women with risk factors. J Matern Fetal Neonatal Med. 2015;28(2):153-7.	Population	Asymptomatic population

Table 53 Parsagen (PartoSure) Submitted citations

No.	Citation	Reason for Exclusion	Further detail
1.	Bolotskikh V, Borisova V. Combined value of placental alpha microglobulin-1 detection and cervical length via transvaginal ultrasound in the diagnosis of preterm labor in symptomatic patients. J Obstet Gynaecol Res. 2017.	Included	
2.	Echebiri NC, McDoom MM, Aalto M, Pullen J, Doyle NM. Placental alpha-microglobulin-1 and combined traditional diagnostic test: a cost-benefit analysis. Obstetrics & Gynecology. 2014;123:3S-S.	Population	Ruptured membranes
3.	Echebiri NC, McDoom MM, Pullen JA, Aalto MM, Patel NN, Doyle NM. Placental alpha-microglobulin-1 and combined traditional diagnostic test: a cost-benefit analysis. American Journal of Obstetrics & Gynecology. 2015;212(1):77.e1-.e10.	Population	Ruptured membranes
4.	Fatkullin I, et al. Utilization of a novel biomarker test (PARTOSURE PAMG-1) to reduce the length of stay in patients with threatened preterm labor and a short cervix. Journal of Maternal-Fetal & Neonatal Medicine. 2016;29(S1):283.	Abstract	Not enough information
5.	Hadzi-Lega M et al. Comparison of PAMG-1 and phIGFBP-1 Tests for the Prediction of Preterm Delivery in Patients with Preterm Labor. Open Journal od Obstetrics and Gynecology. 2017;7:358-68.	Included	
6.	Heverhagen A. Placental Alpha Microglobulin-1 In Combination With Transvaginal Ultrasound For Prediction Of Preterm Birth. Journal of Perinatal Medicine. 2015;43(S1):240.	Abstract	Not enough information
7.	Konoplyannikov A et al. PAMG-1 biomarker test (PARTOSURE) in combination with transvaginal ultrasound for improved assessment of spontaneous preterm birth in patients with threatened preterm labor. Journal of Maternal-Fetal & Neonatal Medicine. 2016;29(S1):278.	Abstract	Not enough information
8.	Lotfi G, Faraz S, Al Swalhee N, Nasir R, Somini S, Abdeldayem R, et al. Evaulation of PAMG-1 for the prediction of preterm birth in patients symptomatic of preterm labour. Journal of Perinatal Medicine. 2015;43((S1)):250.	Abstract	Not enough information
9.	Lou et al. Is PartoSure effective in assessing preterm birth? BJOG. 2016;123(S2):89.	Abstract	Not enough information
10.	M. Ravi et al. Evaluation Of The Quantitative Fetal Fibronectin Test And Partosure™ (Placental Alpha Microglobulin-1 [Pamg-1]) For The Prediction Of Spontaneous Preterm Birth (Sptb) In Patients With Signs And Symptoms Suggestive Of Preterm Labor. Journal of pediatric and neonatal individualized medicine. 2017;6(1):ABS 50.	Abstract	Not enough information
11.	Nikolova T, Bayev O, Nikolova N, Di Renzo GC. Evaluation of a novel placental alpha microglobulin-1 (PAMG-1) test to predict spontaneous preterm delivery. J Perinat Med. 2014;42(4):473-7.	Included	
12.	Nikolova T, Bayev O, Nikolova N, Di Renzo GC. Comparison of a novel test for placental alpha microglobulin-1 with fetal fibronectin and cervical length measurement for the prediction of imminent spontaneous preterm delivery in patients with threatened preterm labor. J Perinat Med. 2015;43(4):395-402.	Included	
13.	Nikolova T, Uotila J, Nikolova N, Borisova VY, Bolotskikh VM. 16: Do PAMG-1 or phIGFBP-1 biomarkers improve the prediction of imminent spontaneous preterm delivery in PTL symptomatic women with non-obvious cervical length (CL)? American Journal of Obstetrics & Gynecology.216(1):S11-S2.	Abstract	Not enough information
14.	Melchor J et al. Retrospective Analysis On The Efficacy Of The Pamg-1 Test And The Fetal Fibronectin Test In Assessing Preterm Birth In Symptomatic Women Attending An Emergency Obstetric Unit. Conference: 1st World Congress on Maternal Fetal and Neonatal Medicine.Research Institute: Cruces University Hospital. Vizcaya. Spain. BioCruces Health Research Institute.	Abstract	Not enough information
15.	Van Holsbeke et al. Comparison of the fetal fibronectin (Rapid fFN) and placental alpha microglobulin-1 (PartoSure) tests for predicting imminent spontaneous preterm birth. Ultrasound in Obsetrics & Gynecology. 2016;48(S1):84.	Abstract	Not enough information
16.	Wing D et al. PAMG-1 (PARTOSURE™) vs. fFN to Assess Risk of Preterm Delivery in Symptomatic Women. Conference: KU Medical Centre/ UC Irvine Health Institute.	Abstract	Not enough information

17	PartoSure GA Analysis	Table	Not enough information

Table 54 Studies excluded at Full Text with reasons

No.	Citation	Reason for Exclusion
1	Caroline VH, Annick C. Comparison of the fetal fibronectin (rapid FFN) and placental alpha microglobulin-1 (partosure) tests for predicting imminent spontaneous preterm birth in patients with threatened preterm labor. Journal of Perinatal Medicine Conference: 12th World Congress of Perinatal Medicine. 2015;43(no pagination).	Abstract
2	Desjardins PR, Dansereau J, Hoag GN. Comparing the clinical effectiveness of Fetal Fibronectin and IGFBP-1 measurements in cervico-vaginal secretions, in predicting preterm deliveries. Clinical Chemistry. 2008;54(6):A39-A40.	Abstract
3	Ehsanipoor RM, Swank M, Jwa SC, Wing DA, Tarabulsi G, Blakemore KJ. Placental alpha-microglobulin-1 in vaginal secretions as a predictor of preterm birth in women with evidence of preterm labor. Reproductive Sciences. 2014;1):155A.	Abstract
4	Fatkullin I, Akhmetgaliev A, Matveeva E, Seeger S. Utilization of a novel biomarker test (PARTOSURE PAMG-1) to reduce the length of stay in patients with threatened preterm labor and a short cervix. Journal of Maternal-Fetal and Neonatal Medicine. 2016;29:283.	Abstract
5	Grobman W, Welshman E, Calhoun E. Does fetal fibronectin use in the diagnosis of preterm labor affect physician behavior and health care costs? A randomized Trial. American journal of obstetrics and gynecology. 2002;187(6 Supplement):S80.	Abstract
6	Grobman WA, Welshman EE, Calhoun EA, Ramsey PS. Fetal fibronectin results did not reduce medical resource use for women with preterm uterine contractions. Evidence-based Obstetrics and Gynecology. 2005;7(3):118-9.	Abstract
7	Hansen W, Lowe M, Zimmerman B. Effect of the fetal fibronectin assay on preterm labor management. American journal of obstetrics and gynecology [Internet]. 2001; 185(6 Suppl):[S136 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/596/CN-00387596/frame.html.	Abstract
8	Heverhagen A. Placental alphamicroglobul in-1 in combination with transvaginal ultrasound for prediction of preterm birth. Journal of Perinatal Medicine Conference: 12th World Congress of Perinatal Medicine. 2015;43(no pagination).	Abstract
9	Heverhagen A, Baumann M, Raio L, Surbek D. Placental alpha-microglobulin-1 in combination with transvaginal ultrasound for prediction of preterm birth. American journal of obstetrics and gynecology. 2015;212(1, Suppl. S):S81.	Abstract
10	Heverhagen A, Muller M, Schleussner E, Deruelle P, Raio L, Surbek D. The prediction of preterm birth using placental alpha-microglobulin-1 in combination with transvaginal ultrasound. Reproductive Sciences. 2016;1):131A-2A.	Abstract
11	Hillman-Cooper C, Ghag K, Dempsey A, Denbow M, Lopez Bernal A. Actim partus-the first year at St. Michael's Hospital, Bristol. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2014;99:A158-A9.	Abstract
12	Holmgren C, Lacoursiere DY, Esplin MS. Clinical predictors of a false negative fetal fibronectin (FFN). American journal of obstetrics and gynecology. 2007;197(6):S204-S.	Abstract
13	Kang JH, Lee SE, Park C-W, Jun JK, Romero R, Yoon BH. Cervical fetal fibronectin: An index of intra-amniotic inflammation, histologic chorioamnionitis and impending preterm delivery in patients with preterm labor and intact membranes. American journal of obstetrics and gynecology. 2007;197(6, Suppl. S):S47.	Abstract
14	Karunakaran B, Berry J, Parasuraman R. Can we raise the threshold for negative fetal fibronectin result? Archives of Disease in Childhood: Fetal and Neonatal Edition. 2012;97:A89.	Abstract
15	Konoplyannikov A, Lysyuk I, Sokolyan A, Pipia N, Apresyan S, Karasova A. PAMG-1 biomarker test (PARTOSURE) in combination with transvaginal ultrasound for improved assessment of spontaneous preterm birth in patients with threatened preterm labor. Journal of Maternal-Fetal and Neonatal Medicine Conference: 25th European Congress of Perinatal Medicine Netherlands Conference Start: 20160615 Conference End: 20160618 [Internet]. 2016; 29:[278 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/387/CN-01214387/frame.html.	Abstract
16	Kuhnert M. Individual management of preterm labour by using rapid fetal fibronectintesta. Journal of Maternal-Fetal and Neonatal Medicine. 2014;27:392-3.	Abstract

No.	Citation	Reason for Exclusion
17	Kung RWK, Northridge R, Nicoll AE. Fetal fibronectin (FFN) and iatrogenic pre-term birth. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2012;97:A80-A1.	Abstract
18	Lawin-O'Brien A, Jesner O, Biswas C. Evaluation of fetal fibronectin testing, subsequent management and outcomes in a London teaching hospital. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2014;99:A119.	Abstract
19	Lega MH. Prediction of preterm delivery in patients in preterm labor. Journal of Perinatal Medicine Conference: 12th World Congress of Perinatal Medicine. 2015;43(no pagination).	Abstract
20	Lotfi G, Faraz S, Al Swalhee N, Nasir R, Somini S, Abdeldayem R, et al. Evaluation of PAMG-1 for the prediction of preterm birth in patients symptomatic of preterm labor. Journal of Perinatal Medicine Conference: 12th World Congress of Perinatal Medicine. 2015;43(no pagination).	Abstract
21	Lou YY, Ajay B. Is PartoSure effective in assessing preterm birth? Bjog-an International Journal of Obstetrics and Gynaecology. 2016;123:89-.	Abstract
22	Mahavarker S, Osman MW, Ten-Hof J. Phosphorylated insulin-like growth factor binding protein-1 (ACTIM PARTUS) testing in management of threatened preterm labour. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2012;97:A99-A100.	Abstract
23	Nikolova N, Nikolova T, Jovchevski S, Micevska M. Phosphorylated insulin-like growth factor binding protein-1 in the prediction of preterm delivery in patients with preterm labor. Journal of Maternal-Fetal and Neonatal Medicine. 2016;29:56.	Abstract
24	Nikolova N, Nikolova T, Jovchevski S, Micevska M, Nikolovski S. A comparison between the phosphorylated insulin-like growth factor binding protein-1 and the cervical length in prediction of sampling to delivery time in patients with preterm labor. Journal of Maternal-Fetal and Neonatal Medicine. 2016;29:56.	Abstract
25	Nikolova T, Uotila J, Nikolova N, Borisova V, Bolotskikh V. Do PAMG-1 or phIGFBP-1 biomarkers improve the prediction of imminent spontaneous preterm delivery in PTL symptomatic women with non-obvious cervical length (CL)? American journal of obstetrics and gynecology Conference: 37th annual meeting of the society for maternal-fetal medicine: the pregnancy meeting United states Conference start: 20170123 Conference end: 20170128 [Internet]. 2017; 216(1 Supplement 1):[S11-s2 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/156/CN-01304156/frame.html.	Abstract
26	Northridge R, Liu H, Youssef R, Nicoll AE. Predicting pre-term birth using fetal fibronectin (FFN) in Ninewells Hospital, Dundee. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2011;96:Fa69.	Abstract
27	Ponnusamy V, Farrer K. Impact of antenatal tests on acute in-utero transfers. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2012;97:A87.	Abstract
28	Radford S, Costa F, Junior E, Sheehan P. Clinical application of quantitative fetal fibronectin for the prediction of preterm birth in symptomatic women. BJOG: An International Journal of Obstetrics and Gynaecology Conference: RCOG World Congress 2016 United Kingdom Conference Start: 20160620 Conference End: 20160622 [Internet]. 2016; 123:[158-9 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/043/CN-01179043/frame.html.	Abstract
29	Randhawa TS, Kwek K. Assessment of the predictive value of phosphorylated insulinlike growth factor binding protein-1 for preterm delivery. Journal of Maternal-Fetal and Neonatal Medicine. 2010;23:294.	Abstract
30	Spyroulis C, Tsitlakidis C. Prospective study of the Actim Partus test to predict preterm delivery between 24-34 weeks of gestation in Glan Clwyd Hospital. Period of study between August 2014-February 2015. BJOG: An International Journal of Obstetrics and Gynaecology. 2016;123:8.	Abstract
31	Thandayathany V, Yassin MAJM, Omar MH, Ismail NAM, Tamil AM, Kampan NC. Fetal fibronectin rapid test versus phosphorylated insulin-like growth factor-1 (phIGFBP-1) as bedside test kits for prediction of preterm delivery in the clinical setting. BJOG: An International Journal of Obstetrics and Gynaecology. 2012;119:5.	Abstract
32	Van Holsbeke C, Dam K, Staelens A, Mesens T, Corremans A. OP11.04: Comparison of the fetal fibronectin (Rapid fFN) and placental alpha microglobulin-1 (PartoSure) tests for predicting imminent spontaneous preterm birth. Ultrasound in Obstetrics & Gynecology. 2016;48:84-.	Abstract
33	Sophie W, Tiphaine B, Celine C, Aurelie DB. Preterm labor: Reliability of the partosure test before and after digital and transvaginal ultrasound examinations. Journal of Perinatal Medicine Conference: 12th World Congress of Perinatal Medicine. 2015;43(no pagination).	Abstract

No.	Citation	Reason for Exclusion
34	Bare T, Mitre A, Shahinaj R, Bakiri F. Fetal fibronectin and cericometry in patients at increased risk for premature birth. Journal of Perinatal Medicine Conference: 12th World Congress of Perinatal Medicine. 2015;43(no pagination).	Abstract
35	Benson J, Ghidini A, Landy H, Gilo N, Drassinower D, Poggi S. Fetal fibronectin for evaluation of preterm labor in the setting of cervical cerclage. American journal of obstetrics and gynecology. 2009;201(6):S68-S9.	Abstract
36	Berghella V, Hayes E, Visintine J, Baxter J. FETAL FIBRONECTIN TESTING FOR REDUCING THE RISK OF PRETERM BIRTH. American journal of obstetrics and gynecology. 2008;199(6):S214-S.	Abstract
37	Burrus DR, Ernest JM, Veille JC. Efficacy of four markers of preterm labor (PTL) at predicting preterm delivery (PTD) in a population presenting with contractions and cervical change. American journal of obstetrics and gynecology. 1995;172(1 PART 2):405.	Abstract
38	Burrus DR, Ernest JM, Veille JC. Efficacy of four markers of preterm labor (PTL) at predicting latency (LAT) in a population receiving tocolytic therapy. American journal of obstetrics and gynecology. 1995;172(1 PART 2):405.	Abstract
39	Goepfert AR, Network NM. The preterm prediction study: Quantitative fetal fibronectin (FFN) values and the prediction of spontaneous preterm birth (SPTB). American journal of obstetrics and gynecology. 1997;176(1 PART 2):S52.	Abstract
40	Goldenberg RL. Vaginal fetal fibronectin (V-fFN) levels at 8-22 weeks and subsequent spontaneous preterm birth (SPB). American journal of obstetrics and gynecology. 2000;182(1 Part 2):S32.	Abstract
41	Greenhagen JB, Van Wagoner J, Dudley D, Hunter C, Mitchell M, Casal D, et al. The value of fetal fibronectin as a predictor of preterm delivery in low risk women. American journal of obstetrics and gynecology. 1996;174(1 PART 2):471.	Abstract
42	Grover CM. Use of the fetal fibronectin assay to avert unnecessary hospitalization and bedrest in patients with preterm labor symptoms: a cost analysis. Primary Care Update for Ob/Gyns. 1998;5(4):178-9.	Abstract
43	Hadzi Lega M, Daneva Markova A, Micevska M, Atanasovska Bosku A, Gjirevski V. Cervical length and phosphorilated insulin like growth factor binding protein-1 as a predictors of spontaneus preterm delivery in symptomatic women. Journal of Maternal-Fetal and Neonatal Medicine. 2014;27:399.	Abstract
44	Hedriana HL, Bliss MC, Gilbert WM. Implementation of fetal fibronectin (FFN) testing in a large private practice hospital is safe and effective for the diagnosis and management of preterm labor (PTL). American journal of obstetrics and gynecology. 2005;193(6):S52-S.	Abstract
45	Henrich W. Cervicometry and fibronectin role. Journal of Maternal-Fetal and Neonatal Medicine. 2010;23:39.	Abstract
46	Hillman-Cooper C, Denbow M, Lopez Bernal A. Identification and quantification of maternal plasma proteins that predict preterm labour. BJOG: An International Journal of Obstetrics and Gynaecology. 2013;120:160-1.	Abstract
47	Hillman-Cooper CS, Denbow M, Bernal AL. Identification and quantification of maternal plasma proteins that predict preterm labour. Endocrine Reviews Conference: 95th Annual Meeting and Expo of the Endocrine Society, ENDO. 2013;34(3 SUPPL. 1).	Abstract
48	Hong J-S, Yoon BH, Romero R, Shim S-S, Sohn Y-K, Park C-W, et al. Patients with a positive cervical fetal fibronectin followed by a negative test have a reduced likelihood of spontaneous preterm delivery. American journal of obstetrics and gynecology. 2003;189(6 Supplement):S173.	Abstract
49	Iriye B, Hancock J, Wold S, Huang W, Gorski L, Hancock L. Fetal fibronectin (fFN) testing in symptomatic patients for preterm labor and McDonald cerclage. American journal of obstetrics and gynecology. 2004;191(6):S111-S.	Abstract
50	Joffe G. The use & cost benefit of fetal fibronectin testing in an integrated health care system. Clinical Chemistry. 1999;45(6):S32-S3.	Abstract
51	Kay S, McKie L, Ajoku K, Roberts N. Fetal fibronectin testing in a district general hospital setting - Are we changing our management of women with symptomatic preterm labour? BJOG: An International Journal of Obstetrics and Gynaecology. 2015;122:186.	Abstract
52	Khizar SBK, Vallabu S. How to introduce a new service in clinical practice. BJOG: An International Journal of Obstetrics and Gynaecology. 2013;120:96.	Abstract
53	Kiefer D, Keeler S, Rust O, Demishev M, Muscat J, Bornstein E, et al. Is fetal fibronectin (FFN) a marker of intra-amniotic inflammation in patients with midtrimester short cervix? American journal of obstetrics and gynecology. 2009;201(6):S60-S.	Abstract

No.	Citation	Reason for Exclusion
54	Knee A, Belisle E, Markenson G, Malshe A, Plevyak M, Bsat F, et al. Do pregnancies complicated by preterm births with a negative fetal fibronectin screen have different characteristics compared to those that have positive fibronectin results? American journal of obstetrics and gynecology. 2011;204:S190-S.	Abstract
55	Knight B, Smith H, Tregarthen A, Lapointe J, Shennan A, Taylor MJO, et al. Quantitative fetal fibronectin as a predictor of time to delivery in women undergoing induction of labour: a pilot study. Bjog-an International Journal of Obstetrics and Gynaecology. 2016;123(13):E5-E.	Abstract
56	Kreaden U, Sheldon L, Lockhart T, Paulson M, Shorter S, Martinez R, et al. Natural distribution of fetal fibronectin (fFN) among symptomatic and asymptomatic pregnant women. Clinical Chemistry. 1999;45(6):A86-A7.	Abstract
57	Krems J, McGregor J, Hastings C, Hanson J. Fetal fibronectin in cervical secretions: An aid for predicting delivery in women at high risk from preterm delivery. American journal of obstetrics and gynecology. 1995;172(1 PART 2):406.	Abstract
58	Kuhrt K, Seed P, Smout E, Hezelgrave N, Shennan A. The development and validation of a new tool to predict spontaneous preterm birth in high-risk women using quantitative fetal fibronectin and cervical length. Bjog-an International Journal of Obstetrics and Gynaecology. 2013;120:14-5.	Abstract
59	Kurtzman J, Chandiramani M, Riley A, Poston L, Das A, Shennan A. Quantitative fetal fibronectin screening at 24 weeks delineates the risk of spontaneous preterm delivery in asymptomatic patients with a prior history of preterm premature rupture of membranes (PPROM). American journal of obstetrics and gynecology. 2009;1):S227.	Abstract
60	Kurtzman J, Jimenez-Solis G, Rosalesortiz S, Martinez-Alvarez O, Das A, Antonio AMJ. CERVICAL LENGTH SIGNIFICANTLY MODIFIES THE RISK OF PRETERM DELIVERY IN LOW RISK PATIENTS WITH SYMPTOMATIC PRETERM LABOR AND A POSITIVE FETAL FIBRONECTIN. American journal of obstetrics and gynecology. 2008;199(6):S212-S.	Abstract
61	Levine LD, Romero JA, Pappas H, Weber AL, Elovitz MA. A prospective blinded cohort study evaluating fetal fibronectin and cervical length in predicting preterm birth in symptomatic women. American journal of obstetrics and gynecology. 2016;214(1):S441-S2.	Abstract
62	Luzzi V, Hankins K, Gronowski AM. Accuracy of the rapid fetal fibronectin TLi system in predicting preterm delivery. Clinical Chemistry. 2003;49(3):501-2.	Abstract
63	Maharajan P, Elliot J, Thampi P. Efficacy of fibronectin and its impact on the use of atosiban-cost implication. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2012;97:A92.	Abstract
64	Mancuso M, Biggio J, Owen J. Fetal fibronectin and latency period in women with arrested preterm labor. American journal of obstetrics and gynecology. 2013;208(1):S221-S.	Abstract
65	Mateus J, Pereira L, Baxter J, Berghella V, Cheku B, Tolosa JE. Fetal fibronectin compared to cervical dilation in clinical practice. American journal of obstetrics and gynecology. 2003;189(6):S165-S.	Abstract
66	McKenna DS, Chung K, Shubert P, Iams JD. The effect of digital examination on the expression of fetal fibronectin. American journal of obstetrics and gynecology. 1998;178(1 PART 2):S188.	Abstract
67	Merriam A, Sciscione A, Daniel L. A prospective evaluation of the effectiveness of vaginal fetal fibronectin to predict preterm birth in women with symptoms of preterm labor in the outpatient setting. American journal of obstetrics and gynecology. 2014;210(1):S224-S.	Abstract
68	Mogami H, Kondo E, Fujita K, Chigusa Y, Nishimura F, Ujita M, et al. FETAL FIBRONECTIN IS A PATHOGENIC FACTOR OF PRETERM PROM AND PRETERM BIRTH. Placenta. 2013;34(10):A4-A.	Abstract
69	Morrison JC, Naefi RWI, Botti JJ, Katz M, Belluomini JM. Uterine activity and fetal fibronectin are independent and combined predictors of spontaneous preterm birth. American journal of obstetrics and gynecology. 1995;172(1 PART 2):406.	Abstract
70	Muller A, McCullough M, Obican S, Norton H, Carter J, Gonzalez-Quintero VH. "false" positive fetal fibronectin and rates of adverse obstetrical outcomes. American journal of obstetrics and gynecology. 2006;195(6):S233-S.	Abstract
71	Muzaffar S, Behrens R, Barns F, Walker S. Quantitative fetal fibronectin and preterm delivery. Bjog-an International Journal of Obstetrics and Gynaecology. 2013;120:595-6.	Abstract

No.	Citation	Reason for Exclusion
72	Myint H, Singh V, Roberts NJ. Introducing fetal fibronectin testing to a District General Hospital in women with symptoms of preterm labour. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2011;96:Fa83.	Abstract
73	Myint HH, Ciopec A, Roberts N, Sinha P. Use of fetal fibronectin testing in a district general hospital in women with symptoms of preterm labour. Bjog-an International Journal of Obstetrics and Gynaecology. 2013;120:41-2.	Abstract
74	Ness A, Airoldi J, Modena A, Baxter J, Berghella V. Triage of women with possible preterm labor using transvaginal ultrasound cervical length and fetal fibronectin. American journal of obstetrics and gynecology. 2005;193(6):S58-S.	Abstract
75	O'Brien J, Mercer B, Pitts DW, Sibai BM. A comparison of fetal fibronectin, prolactin, and cervical examination for the prediction of preterm delivery in symptomatic patients. American journal of obstetrics and gynecology. 1995;172(1 PART 2):406.	Abstract
76	O'Byrne L, Daly S. Audit of fetal fibronectin use in a tertiary referral hospital in the triaging of threatened preterm labour. BJOG: An International Journal of Obstetrics and Gynaecology. 2016;123:13-4.	Abstract
77	Osorio M, Neiva R, Montes L, Silva J, Pinelo S, Pinho M, et al. The impact of fetal fibronectin assay on preterm labor management: a randomized controlled trial. Puertas A, Montoya F, Romero J, Hurtado J, Manzanares S, editors2010. 535-8 p.	Abstract
78	Osorio M, Neiva R, Montes L, Silva J, Pinelo S, Pinho M, et al. The impact of fetal fibronectin assay on preterm labor management: A randomized controlled trial. Journal of Maternal-Fetal and Neonatal Medicine. 2010;23:304-5.	Abstract
79	Peaceman AM, Andrews WW, Thorp JM, Cliver SP, Lukes A, Group FFS. Fetal fibronectin as a predictor of preterm birth in symptomatic patients: A multicenter trial. American journal of obstetrics and gynecology. 1996;174(1 PART 2):303.	Abstract
80	Pilpel S, Markenson G, Hoffman D. A comparison of fetal fibronectin testing in predicting preterm birth in twin verses singleton gestations. American journal of obstetrics and gynecology. 2002;187(6 Supplement):S120.	Abstract
81	Pradhan-Drews M, Drews F, Goddard R. Audit on preterm labour. BJOG: An International Journal of Obstetrics and Gynaecology. 2013;120:431-2.	Abstract
82	Radford S, Abbott D, Seed P, Kemp J, Shennan A. Quantitative fetal fibronectin for the prediction of preterm birth in symptomatic women. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2012;97:A73.	Abstract
83	Radford S, Smith J, Shennan A, Wallace EM. Interhospital transfers of women in threatened preterm labour to Victorian level 6 hospitals. Journal of Paediatrics and Child Health. 2015;51:51.	Abstract
84	Ray D, Dyson D. Fetal fibronectin as a predictor of preterm delivery in symptomatic patients. American journal of obstetrics and gynecology. 2001;184(1):S37.	Abstract
85	Rebarber A, Sfakianaki AK, Kuczynski E, Paidas M, Russo K, Maturi J, et al. Fetal fibronectin is a valid predictor of spontaneous preterm delivery in patients with cervical cerclages. American journal of obstetrics and gynecology. 2001;185(6 Supplement):S128.	Abstract
86	Revah A, Sue-A-Quan A, Hannah M. Fetal fibronectin as a predictor of preterm birth: A systematic review of the literature. American journal of obstetrics and gynecology. 1997;176(1 PART 2):S53.	Abstract
87	Richards KM, Troeger KA, Petrilla A, Agatep B, Johnsrud M, Blackwell SC. Burden of preterm labor in the texas medicaid population. Value in Health. 2016;19 (3):A178.	Abstract
88	Rizzo G, Capponi A, Arduini D, Romanini C. The prognostic value of interleukin-8 and fetal fibronectin concentrations in cervical secretions in patients with preterm labor. American journal of obstetrics and gynecology. 1997;176(1 PART 2):S6.	Abstract
89	Roman AS, Koklanaris N, Roshan D, Paidas M, Mulholland J, Levitz M, et al. "Blind" vaginal fetal fibronectin as a predictor of spontaneous preterm delivery: A novel sampling technique. Journal of the Society for Gynecologic Investigation. 2004;11(2):204A-5A.	Abstract
90	Rosenblatt KP, Nicolaides KH, Romero R, Bryant-Greenwood PK, Syngelaki A, Menon R, et al. Proteomic derived biomarker panel for first trimester identification of patients at risk for pre-term delivery. Reproductive Sciences. 2013;1):80A.	Abstract
91	Salari K, Treadwell M, Thompson K. Predictors of preterm birth: a comparison of cervical length, fetal fibronectin and fetal adrenal gland enlargement. American journal of obstetrics and gynecology. 2015;212(1):S179-S80.	Abstract

No.	Citation	Reason for Exclusion
92	Sasson I, Triche EW, Azpurua H, Romano M, Paidas M. SHORT CERVIX AND NEGATIVE FETAL FIBRONECTIN: LOW RISK OF DELIVERY WITHIN 2 WEEKS OF SCREENING. American journal of obstetrics and gynecology. 2008;199(6):S49-S.	Abstract
93	Schmiedel D, Fischer T, Von Tucher E, Henrich W, Slowinski T, Thomas A. Combination of cervical sonoelastography with cervical length measurement and fetal fibronectin to improve prediction of preterm birth. Ultraschall in der Medizin, Supplement Conference: Ultraschall. 2013;34(no pagination).	Abstract
94	Simhan H. Comparison of self-collected vs speculum-collected measurement of quantitative fetal fibronectin. American journal of obstetrics and gynecology. 2016;214(1):S310-S1.	Abstract
95	Stern V, Amabebe E, Anumba D. The association between cervicovaginal fetal fibronectin, the metabolic markers of the vaginal microbiome, and preterm birth. Bjog-an International Journal of Obstetrics and Gynaecology. 2016;123:79-80.	Abstract
96	Stern V, Anumba D. Screening for preterm birth with serial cervical length measurement and fetal fibronectin quantification - does rate of change predict early delivery? Bjog-an International Journal of Obstetrics and Gynaecology. 2016;123:80-.	Abstract
97	Stern V, Anumba DOC. Predicting Preterm Birth - The Performance of Ultrasound Cervical Length, Quantitative Fetal Fibronectin and Vaginal Fluid pH Differs by Risk of Preterm Birth and Symptoms of Preterm Labour. Reproductive Sciences. 2016;23:106A-A.	Abstract
98	Stern V, Healey J, Brown B, Anumba D. Assessing the risk of spontaneous premature birth by electrical impedance spectroscopy of the cervix: The ecclippx study. Reproductive Sciences. 2015;22:331A.	Abstract
99	Stern V, Healey J, Jones G, Dixon S, Walters S, Brown B, et al. The ECCLIPPx study: Electrical impedance prediction of preterm birth. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2014;99:A113-A6.	Abstract
100	Swamy G, Simhan H, Seglin H, Heine P. The clinical utility of fetal fibronectin. American journal of obstetrics and gynecology. 2001;185(6 Supplement):S136.	Abstract
101	Swancoat S. "Recent Vaginal Exam" No Longer a Contraindication to Fetal Fibronectin Sampling: A Prospective Study. Obstetrics and Gynecology. 2016;127:16S-S.	Abstract
102	Thompson KG, Vaillancourt C, Moore S. Air medical transfer of patients in preterm labour: current treatments and tocolytics provided in Ontario. Canadian Journal of Emergency Medicine. 2013;15:S6.	Abstract
103	Turitz AL, Ananth C, Gyamfi-Bannerman C. A comparison of quantitative pre- and post-vaginal manipulation fetal fibronectin. American journal of obstetrics and gynecology. 2017;216 (1 Supplement 1):S197.	Abstract
104	van Baaren GJ, Bruijn M, Vis J, Wilms F, Oudijk M, Kwee A, et al. False-positive, false-negative and uninterpretable results in fetal fibronectin testing during the APOSTEL1 study; which factors do contribute? American journal of obstetrics and gynecology. 2014;210(1):S395-S.	Abstract
105	van der Krogt L, Hezelgrave N, Seed P, Shennan A. Prediction of spontaneous preterm birth using fetal fibronectin in women with a low-lying placenta. Bjog-an International Journal of Obstetrics and Gynaecology. 2016;123:85-.	Abstract
106	Vanner T, Rooney M, Blackburn S. Trends in In-utero transfer: Royal Wolverhampton Hospitals NHS Trust 2007 to 2010. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2011;96:Fa38.	Abstract
107	Watson H, Carter J, Tribe RM, Shennan AH. The QUIPP app: A safe alternative to a treat-all strategy for threatened preterm labour. BJOG: An International Journal of Obstetrics and Gynaecology. 2017;124:30-1.	Abstract
108	Watson HA, Ridout A, Ross G, Shennan AH. Decision-making about preterm birth using the QUIPP app: A survey of women's experiences. BJOG: An International Journal of Obstetrics and Gynaecology. 2017;124:132.	Abstract
109	Akercan F, Cirpan T, Kazandi M, Terek MC, Mgoyi L, Ozkinay E. The value of the insulin-like growth factor binding protein-1 in the cervical-vaginal secretion detected by immunochromatographic dipstick test in the prediction of delivery in women with clinically unconfirmed preterm premature rupture of membranes. European Journal of Obstetrics Gynecology and Reproductive Biology. 2005;121(2):159-63.	Comparator
110	Akercan F, Kazandi M, Sendag F, Cirpan T, Mgoyi L, Terek MC, et al. Value of cervical phosphorylated insulinlike growth factor binding protein-1 in the prediction of preterm labor. Journal of Reproductive Medicine. 2004;49(5):368-72.	Comparator

No.	Citation	Reason for Exclusion
111	Gibson S, Hezelgrave NL, Shennan AH. The impact of quantitative fetal fibronectin on the clinical management of women symptomatic of preterm labour. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2014;99:A151-A5.	Comparator
112	Kuhrt K, Hezelgrave N, Foster C, Seed PT, Shennan AH. Development and validation of a tool incorporating quantitative fetal fibronectin to predict spontaneous preterm birth in symptomatic women. Ultrasound in Obstetrics & Gynecology. 2016;47(2):210-6.	Comparator
113	Cervicovaginal fetal fibronectin test for predicting risk of spontaneous preterm birth. ACOG Clinical Review. 2002;7(10):3-4.	Design
114	Management of preterm labor. International Journal of Gynecology & Obstetrics. 2003;82(1):127-35.	Design
115	Erratum. Journal of Maternal-Fetal & Neonatal Medicine. 2016;29(19):3212-.	Design
116	Abbott D, Radford S, Foster C, Vousden N, Shennan A. Longitudinal trend of quantitative fetal fibronectin in the prediction of delivery following insertion of a rescue cerclage. Journal of Obstetrics & Gynaecology. 2013;33(4):414-5.	Design
117	Adams TM, Kinzler WL, Chavez MR, Fazzari MJ, Vintzileos AM. Practice patterns in the timing of antenatal corticosteroids for fetal lung maturity. Journal of Maternal-Fetal & Neonatal Medicine. 2015;28(13):1598-601.	Design
118	Afify E, Pasupathy D, Briley A, Seed P, Poston L, Shennan A. The effect of maternal obesity on preterm delivery in a high risk cohort. American journal of obstetrics and gynecology. 2011;204 (1 SUPPL.):S224.	Design
119	Barnard EP. Fetal Fibronectin Testing for Prevention of Preterm Birth in Singleton Pregnancies With Threatened Preterm Labor: A Systematic Review and Metaanalysis of Randomized Controlled Trials. Obstetrics and Gynecology. 2016;128(2):404-.	Design
120	Bernhardt J, Dorman K. Pre-term birth risk assessment tools: exploring fetal fibronectin and cervical length for validating risk. AWHONN Lifelines. 2004;8(1):38-44.	Design
121	Brink S, Chetwynd J, Mannix M. A telling test for early birth... use of a vaginal smear to detect fetal fibronectin. US News & World Report. 1995;119(14):86-.	Design
122	Bruijn MMC, Hermans FJR, Vis JY, Wilms FF, Oudijk MA, Kwee A, et al. Which Factors Contribute to False-Positive, False-Negative, and Invalid Results in Fetal Fibronectin Testing in Women with Symptoms of Preterm Labor? American Journal of Perinatology. 2017;34(3):234-9.	Design
123	Chandiramani M, Di Renzo GC, Gottschalk E, Helmer H, Henrich W, Hoesli I, et al. Fetal fibronectin as a predictor of spontaneous preterm birth: a European perspective. Journal of Maternal-Fetal & Neonatal Medicine. 2011;24(2):330-6.	Design
124	Chien PFW, Khan KS, Ogston S, Owen P. How useful is cervico-vaginal fetal fibronectin in predicting preterm labour - Evidence from an overview. Journal of Obstetrics and Gynaecology. 1998;18(1):95.	Design
125	Colombo DF. Predicting spontaneous preterm birth - Fetal fibronectin and ultrasonography help to rule out labour, not rule it in. British Medical Journal. 2002;325(7359):289-90.	Design
126	Crawshaw SE, Briley AL. Clinical. A new strategy to tackle preterm labour: PREMET trial. RCM Midwives Journal. 1999;2(4):122-4.	Design
127	Daskalakis GJ, Papantoniou NE, Koutsodimas NB, Papapanagiotou A, Antsaklis AJ. Fetal fibronectin as a predictor of preterm birth. Journal of Obstetrics and Gynaecology. 2000;20(4):347-53.	Design
128	Decimoni TC, Sansone D, Etto H, Santos AM, Araujo G, Fonseca ES, et al. SYSTEMATIC LITERATURE REVIEW AND COST-EFFECTIVENESS ANALYSIS OF FETAL FIBRONECTIN TEST FOR PREDICTING PRETERM IN BRAZIL. Value in Health. 2013;16(7):A333-A4.	Design
129	Deshpande S, Asselt A, Tomini F, Armstrong N, Allen A, Noake C, et al. Rapid fetal fibronectin (fFN) testing to predict pre-term birth in women with symptoms of premature labour: A systematic review and cost-effectiveness analysis (Structured abstract). Health Technology Assessment Database [Internet]. 2013; (4):[1 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32012000531/frame.html.	Design

No.	Citation	Reason for Exclusion
130	Elliott J, Miller H, Coleman S, Rhea D, Abril D, Hallbauer K, et al. A randomized multicenter study to determine the efficacy of activity restriction for preterm labor management in patients testing negative for fetal fibronectin. Journal of perinatology [Internet]. 2005; 25(10):[626-30 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/559/CN-00529559/frame.html.	Design
131	Evantash E. Vaginal Fetal Fibronectin to Predict Spontaneous Preterm Birth. JAMA: Journal of the American Medical Association. 2017;318(2):198-9.	Design
132	Faron G, Boulvain M, Irion O. Fetal fibronectin for prediction of preterm delivery: A meta-analysis. American journal of obstetrics and gynecology. 1996;174(1 PART 2):471.	Design
133	Feinberg RF, Kliman HJ. FETAL FIBRONECTIN AND PRETERM LABOR. New England Journal of Medicine. 1992;326(10):708-.	Design
134	French L. Fetal fibronectin to predict preterm delivery. The Journal of family practice. 1998;47(4):250-1.	Design
135	Fuks B, Petrunin DD, Zaraisky EI, Boltovskaya MN, Nazimova SV, Starosvetskaya NA, et al., inventorsDevices and methods for detecting amniotic fluid in vaginal secretions patent US 08114610. 2012 Feb 14 2012.	Design
136	Fuks B, Petrunin DD, Zaraisky EIi, Boltovskaya MN, Nazimova SV, Starosvetskaya NA, et al., inventorsDevices and methods for detecting amniotic fluid in vaginal secretions patent US 09568479. 2017 Feb 14 2017.	Design
137	Gebhardt S, Odendaal HJ. Fetal fibronectin in vaginal secretions--a predictor of preterm delivery? South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde. 1995;85(3):188.	Design
138	Goldenberg RL, Andrews WW, Hauth JC. Markers of preterm birth. Prenatal and Neonatal Medicine. 1998;3(1):43-6.	Design
139	Hee L. Likelihood ratios for the prediction of preterm delivery with biomarkers. Acta Obstetricia Et Gynecologica Scandinavica. 2011;90(11):1189-99.	Design
140	Herbst A, Nilsson C. Diagnosis of early preterm labour. Bjog-an International Journal of Obstetrics and Gynaecology. 2006;113:60-7.	Design
141	Heyborne KD. Fetal fibronectin testing in threatened preterm labor: time for more study! American journal of obstetrics and gynecology. 2017;217(1):94.	Design
142	Hezelgrave NL, Shennan AH, David AL. RATIONAL TESTING Tests to predict imminent delivery in threatened preterm labour. Bmj-British Medical Journal. 2015;350.	Design
143	Jeavons W. Sterile speculum exams & fFN collection. AWHONN lifelines / Association of Women's Health, Obstetric and Neonatal Nurses. 2005;9(3):236-40.	Design
144	Kiss H, Ahner R, Hohlagschwandtner M, Leitich H, Husslein P. Fetal fibronectin as a predictor of term labor: A literature review. Acta Obstetricia Et Gynecologica Scandinavica. 2000;79(1):3-7.	Design
145	Krupa FG, Faltin D, Cecatti JG, Surita FG, Souza JP. Predictors of preterm birth. International Journal of Gynaecology & Obstetrics. 2006;94(1):5-11.	Design
146	Kuin RA, Vis JY, Mol BW. Fetal fibronectin as a short-term predictor of preterm birth in symptomatic patients. Obstetrics & Gynecology. 2010;115(1):186-7.	Design
147	Lamont RF, Dragovic B. A randomised controlled trial of metronidazole for the prevention of preterm birth in women positive for cervicovaginal fetal fibronectin: the PREMET Study by Shennan et al. Bjog-an International Journal of Obstetrics and Gynaecology. 2006;113(7):850-1.	Design
148	McCubbin K, Moore S, MacDonald R, Vaillancourt C. Medical Transfer of Patients in Preterm Labor: Treatments and Tocolytics. Prehospital Emergency Care. 2015;19(1):103-9.	Design
149	Mundy L, Merlin T, Parrella A. A rapid foetal fibronectin assay as a predictive test for women suspected of being in pre-term labour (Structured abstract). Health Technology Assessment Database [Internet]. 2004; (4). Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32006000615/frame.html.	Design

No.	Citation	Reason for Exclusion
150	Musaad SM, Melson LC, Boswell RD. Fetal fibronectin assay may reduce management cost of preterm labour: an interval analysis. Pathology. 2006;38(5):473-4.	Design
151	Nihr H. PartoSure™ Time to Delivery test in suspected premature labour (Structured abstract). Health Technology Assessment Database [Internet]. 2013; (4). Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32013000797/frame.html.	Design
152	O'Sullivan M. Fetal fibronectin as a tool to reduce preterm labour admissions. Australian Midwifery News. 2006;6(1):14-5.	Design
153	Owen P. Fetal fibronectin detection for prediction of preterm birth in low risk women. British Journal of Obstetrics and Gynaecology. 1995;102(12):1019-.	Design
154	Paul RH. Fetal fibronectin and the safety of outpatient management for preterm labor symptoms. American journal of obstetrics and gynecology. 1996;174(2):797-8.	Design
155	Peterson WE, Sprague AE, Reszel J, Walker M, Fell DB, Perkins SL, et al. Women's perspectives of the fetal fibronectin testing process: a qualitative descriptive study. BMC Pregnancy & Childbirth. 2014;14(1):190-.	Design
156	Plaut M, Iams JD. Fetal fibronectin and the safety of outpatient management for preterm labor symptoms [3]. American journal of obstetrics and gynecology. 1996;174(2):797-8.	Design
157	Pucillo K, Munneke S. Fetal fibronectin as predictor of preterm delivery in women with symptoms of preterm labor: women with negative FFN test results are unlikely to deliver in the following 2 weeks; observation is still warranted, but expensive interventions may be unnecessary. Evidence-Based Practice. 2008;11(7):11-2.	Design
158	Rizzo G, Arduini D. Obstetrical markes for very premature delivery. Bevilacqua G, editor2007. 207-13 p.	Design
159	Rizzo G, Capponi A, Angelini E, Muscatello A. Ultrasonographic and biochemical markers of preterm birth. Carrera JM, Cabero L, Baraibar R, editors2001. 608-13 p.	Design
160	Ross GN, Ridout AE, Shennan AH. Optimal clinical risk prediction can be achieved by combining quantitative fetal fibronectin and cervical length, and avoiding thresholds. Acta Obstetricia Et Gynecologica Scandinavica. 2016;95(8):956-.	Design
161	Sanchez-Ramos L. The preterm labor index and fetal fibronectin for prediction of preterm delivery with intact membranes. Obstetrics and Gynecology. 2003;102(1):196-.	Design
162	Schindhelm RK, Hoogenberg J, de Vos MT, Tegelaers FPW. Analytical performance of quantitative fetal fibronectin assay. Ultrasound in Obstetrics & Gynecology. 2016;47(1):127-.	Design
163	Siassakos D, O'Brien K, Draycott T. Healthcare evaluation of the use of atosiban and fibronectin for the management of pre-term labour. Journal of Obstetrics and Gynaecology. 2009;29(6):507-11.	Design
164	Vis JY, Wilms FF, Oudijk MA, Porath MM, Scheepers HCJ, Bloemenkamp KWM, et al. Cost-effectiveness of fibronectin testing in a triage in women with threatened preterm labor: alleviation of pregnancy outcome by suspending tocolysis in early labor (APOSTEL-I trial). Bmc Pregnancy and Childbirth. 2009;9.	Design
165	Vivanti AJ, Benachi A, Rouzier R. "Individualized assessment of preterm birth risk using two modified prediction models" from M. Mailath-Pokorny and colleagues. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2015;188:136-7.	Design
166	White K. Rapid test for fetal fibronectin measures preterm risk for women. Journal of Womens Health. 1998;7(9):1076-.	Design
167	Wiqvist N. FETAL FIBRONECTIN AND PRETERM BIRTH. Acta Obstetricia Et Gynecologica Scandinavica. 1993;72(7):507-8.	Design
168	Bastek JA, Sammel MD, Srinivas SK, McShea MA, Foreman MN, Elovitz MA, et al. Clinical prediction rules for preterm birth can risk stratify patients presenting with preterm labor. American journal of obstetrics and gynecology. 2012;1):S219-S20.	Design
169	Bastek JA, Sammel MD, Srinivas SK, McShea MA, Foreman MN, Elovitz MA, et al. Clinical Prediction Rules for Preterm Birth in Patients Presenting With Preterm Labor. Obstetrics and Gynecology. 2012;119(6):1119-28.	Design

No.	Citation	Reason for Exclusion
170	Tan HZ, Wen SW, Chen XK, Demissie K, Walker M. Early prediction of preterm birth for singleton, twin, and triplet pregnancies. European Journal of Obstetrics Gynecology and Reproductive Biology. 2007;131(2):132-7.	Design
171	Abdelazim IA, Al-Sherbeeny MM, Ibrahim MEM, Fahmy AA, Rabei NH, Aziz Khalifa AA. Insulin-like growth factor binding protein-1/alpha-fetoprotein versus placental alpha microglobulin-1 for diagnosis of premature fetal membranes rupture. Acta Medica International. 2016;3(1):69-74.	Design
172	Akhmetgaliev AR, Fatkullin IF, Munavirova AA, Fatkullin FI. ALGORITHM TO IDENTIFY THE SIGNS OF THREATENED PRETERM LABOUR. Kazanskii Meditsinskii Zhurnal. 2017;98(1):132-6.	Language
173	Benoist G. Prediction of preterm delivery in symptomatic women (preterm labor). Journal De Gynecologie Obstetrique Et Biologie De La Reproduction. 2016;45(10):1346-63.	Language
174	Di Renzo GC, Giardina I, Coata G, Di Tommaso M, Facchinetti F, Petraglia F, et al. [Identification of preterm labor: the role of the fibronectin and ultrasound cervicometry and their association]. Minerva Ginecologica. 2011;63(6):477-83.	Language
175	Florjanski J, Klosek A, Zalewski J, Tomialowicz M, Fuchs T, Heimrath J, et al. [Ultrasonographic assessment of cervical length and measurement of fetal fibronectin level in cervical secretion of pregnant women in prophylaxis of premature deliveries]. Ginekologia Polska. 2001;72(12):1139-43.	Language
176	Hampl M, Friese K, Hofmann I, Melchert F. Quantitative determination of fetal fibronectin in cervical smears: a new marker for evaluating the risk of premature labor. Geburtshilfe und frauenheilkunde [Internet]. 1994; 54(12):[685-90 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/354/CN-00110354/frame.html.	Language
177	Hampl M, Hofmann I, Gallati H, Melchert F, Friese K. FETAL FIBRONECTIN, TUMOR-NECROSIS-FACTOR AND INTERLEUKIN-6 - NEW DIAGNOSIS OF PRETERM DELIVERY AND RUPTURE OF THE MEMBRANES. Archives of Gynecology and Obstetrics. 1993;254(1-4):1470-2.	Language
178	Helmer H, Brunbauer M, Schatten C, Rohrmeister K, Husslein P. Treatment with the oxytocin receptor antagonist atosiban in patients with risk of pre-term delivery. Geburtshilfe Und Frauenheilkunde. 2001;61(10):755-60.	Language
179	Hincz P, Wilczynski J, Pawlowicz P, Borowski D, Krekora M, Szaflik K. [Fetal fibronectin as a predictor of preterm delivery in patients with preterm contractions and cervical changes]. Ginekologia Polska. 2000;71(8):728-32.	Language
180	Lorenz-Eberhardt G, Weiss PAM, Haas J, Puerstner P, Adelwoehrer NE. Premature delivery and fetal fibronectin in the cervical and vaginal secretion. Geburtshilfe und Frauenheilkunde. 1994;54(7):414-6.	Language
181	Mansouri A, Tadjerouni A, El Rabiey G, Baube S, Garnier G, Tribalat S. [Is fetal fibronectin a valid test predictive of premature labor?]. Fertilite Contraception Sexualite. 1997;25(5):380-4.	Language
182	Roubille M, Mailliavin A, Biguet-Vernier B, Bon C, Golfier F, Pichot JC. Fetal fibronectin as predictor of preterm birth. Annales De Biologie Clinique. 1999;57(1):93-7.	Language
183	Surbek D, Boesiger H, Pavic N, Huber P, Almendral AC, Holzgreve W. Fetal fibronectin as a predictor of preterm birth in patients at increased risk. Zeitschrift fuer Geburtshilfe und Neonatologie. 1997;201(1):15-20.	Language
184	Tchobroutsky C. Fetal fibronectin in cervical and vaginal secretions as a predictor of preterm delivery. Contraception Fertilite Sexualite. 1998;26(1):23-5.	Language
185	Velasco JG, Gonzalez AG. FIBRONECTIN AND PREVENTION OF PRETERM DELIVERY. Medicina Clinica. 1995;105(2):54-5.	Language
186	Werlen S, Raia T, Di Bartolomeo A, Chauleur C. Preterm labor: Reproducibility of detection test of PAMG-1 before and after digital examination, and transvaginal ultrasound cervical length. Gynecologie Obstetrique & Fertilite. 2015;43(10):640-5.	Language
187	Wilms FF, van Stralen G, Porath MM, Papatsonis DN, Oei SG, Mol BW, et al. [Predicating imminent preterm labour based on a determination of foetal fibronectin in a vaginal smear]. Nederlands Tijdschrift voor Geneeskunde. 2009;153:B398.	Language

No.	Citation	Reason for Exclusion
188	Abbott DS, Radford SK, Seed PT, Tribe RM, Shennan AH. Evaluation of a quantitative fetal fibronectin test for spontaneous preterm birth in symptomatic women. American journal of obstetrics and gynecology. 2013;208(2).	Outcome
189	Amabebe E, Reynolds S, Stern V, Stafford G, Paley M, Anumba DO. Cervicovaginal Fluid Acetate: A Metabolite Marker of Preterm Birth in Symptomatic Pregnant Women. Frontiers in Medicine. 2016;3:48.	Outcome
190	Asakura H, Fukami T, Kurashina R, Tateyama N, Doi D, Takeshita T. Significance of Cervical Gland Area in Predicting Preterm Birth for Patients with Threatened Preterm Delivery: Comparison with Cervical Length and Fetal Fibronectin. Gynecologic and Obstetric Investigation. 2009;68(1):1-8.	Outcome
191	Ben-Haroush A, Poran E, Yogev Y, Glezerman M. Vaginal fetal fibronectin evaluation before and immediately after ultrasonographic vaginal cervical length measurements in symptomatic women at risk of preterm birth: A pilot study. Journal of Maternal-Fetal & Neonatal Medicine. 2010;23(8):854-6.	Outcome
192	Blackwell S, Shen X, Petrilla AA, Sullivan E, Troeger K. 404: Utilization of fetal fibronectin testing and timing of delivery among emergency department admissions with symptoms of preterm labor. American Journal of Obstetrics & Gynecology. 2017;216:S241-S.	Outcome
193	Bogavac M, Simin N, Ranisavljevic M, Budisic L. The role of insulin-like growth factor in prediction and prevention of preterm delivery. Vojnosanitetski Pregled. 2010;67(11):883-6.	Outcome
194	Bolt LA, Chandiramani M, De Greeff A, Seed P, Shennan AH. Does fetal fibronectin testing change patient management in women at risk of preterm labour? European Journal of Obstetrics & Gynecology and Reproductive Biology. 2009;146(2):180-3.	Outcome
195	Bredaki FE, Lawin-O'Brien A, Jesner O, Forya F, Biswas C, David AL. Quantitative fetal fibronectin to triage women with threatened preterm labour in clinical practice. Bjog-an International Journal of Obstetrics and Gynaecology. 2015;122:104-.	Outcome
196	Browne H, Jassel I, Dhanji A, Bonney E, Simpson N. Use of quantitative fetal fibronectin may improve risk assessment in symptomatic women at risk of preterm birth. Archives of Disease in Childhood: Fetal and Neonatal Edition Conference: 16th Annual Conference of the British Maternal and Fetal Medicine Society Dublin Ireland Conference Publication:. 2013;98(no pagination).	Outcome
197	Chandiramani M, De Greeff A, Filmer J, Bolt LA, Smout E, Shennan AH. The affirm study: Assessment of fetal fibronectin testing to improve preterm management. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2011;96:Fa65-Fa6.	Outcome
198	Echebiri NC, McDoom MM, Aalto M, Pullen J, Doyle NM. Placental alpha-microglobulin-1 and combined traditional diagnostic test: a cost-benefit analysis. Obstetrics & Gynecology. 2014;123:3S-S.	Outcome
199	Elizur SE, Yinon Y, Epatein GS, Seidman DS, Schiff E, Sivan E. Insulin-like growth factor binding protein-1 detection in preterm labor: Evaluation of a bedside test. American Journal of Perinatology. 2005;22(6):305-9.	Outcome
200	Fiorini F, Isted A, Hezelgrave NL, Shennan AH. Prediction of spontaneous preterm birth with quantitative fetal fibronectin in patients with bulging fetal membranes. Bjog-an International Journal of Obstetrics and Gynaecology. 2015;122:123-.	Outcome
201	Fiorini F, Isted A, Hezelgrave NL, Shennan AH. Quantitative fetal fibronectin predicts preterm birth in women with bulging fetal membranes. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2016;203:127-31.	Outcome
202	Gardner M, Rouse D, Joffe G. Cost-benefit analysis of fetal fibronectin screening. American journal of obstetrics and gynecology. 1998;178(1 PART 2):S121.	Outcome
203	Giles W, Knox M, Madsen G, Bisits A, Smith R. The effect of fetal fibronectin usage on admissions to a tertiary maternal fetal medicine unit and cost savings. American journal of obstetrics and gynecology. 1998;178(1 PART 2):S121.	Outcome
204	Goepfert AR, Goldenberg RL, Andrews WW, Hauth JC, Mercer B, Iams J, et al. The Preterm Prediction Study: Association between cervical interleukin 6 concentration and spontaneous preterm birth. American journal of obstetrics and gynecology. 2001;184(3):483-8.	Outcome
205	Gomez R, Romero R, Montiel C, Magendzo A, Pino P, Nien J, et al. Cervico-vaginal fetal fibronectin improves the prediction of preterm delivery estimated by sonographic cervical length in patients with preterm labor and intact membranes. American journal of obstetrics and gynecology. 2002;187(6 Supplement):S80.	Outcome

No.	Citation	Reason for Exclusion
206	Hadzi Lega M, Daneva Markova A, Girevski V. Correlation of cervical length and different biochemical markers in spontaneus preterm birth up to 7 days in symptomatic patients. Journal of Perinatal Medicine Conference: 12th World Congress of Perinatal Medicine. 2015;43(no pagination).	Outcome
207	Hadzi Lega M, Daneva Markova A, Stefanovic M, Tanturovski M. Interleukin 6 and fetal fibronectin as a predictors of preterm delivery in symptomatic patients. Bosnian journal of basic medical sciences / Udruzenje basicnih mediciniskih znanosti = Association of Basic Medical Sciences. 2015;15(1):51-6.	Outcome
208	Hadzi Lega M, Daneva Markova A, Tanturovski M. Interleukin 6 and fetal fibronectin in prediction of preterm delivery in symptomatic patients. Journal of Maternal-Fetal and Neonatal Medicine. 2014;27:373.	Outcome
209	Hadzi-Lega M, Daneva Markova A, Stefanovic M, Tanturovski M. Combination of selected biochemical markers and cervical length in the prediction of impending preterm delivery in symptomatic patients. Clinical and Experimental Obstetrics and Gynecology. 2016;43(1):154-60.	Outcome
210	Hadži-Lega M, Markova AD, Stefanovic M, Tanturovski M. Correlation of cervical length, fetal fibronectin, phIGFBP-1, and cytokines in spontaneous preterm birth up to 14 days from sampling. Journal of Perinatal Medicine. 2015;43(5):545-51.	Outcome
211	Kekki M, Kurki T, Karkkainen T, Hiilesmaa V, Paavonen J, Rutanen EM. Insulin-like growth factor-binding protein-1 in cervical secretion as a predictor of preterm delivery. Acta Obstetricia Et Gynecologica Scandinavica. 2001;80(6):546-51.	Outcome
212	Latifagic A, Balic D, Fatusic Z, Hudic I, Kapidzic M, Habibovic A. Insulin-like growth factor-binding protein-1 (IGFBP-1) in cervical secretions in women with symptoms of preterm delivery. Medicinski Glasnik. 2008;5(2):121-4.	Outcome
213	Paternoster D, Riboni F, Vitulo A, Plebani M, Dell'avanzo M, Battagliarin G, et al. Phosphorylated insulin-like growth factor binding protein-1 in cervical secretions and sonographic cervical length in the prediction of spontaneous preterm delivery. Ultrasound in Obstetrics & Gynecology. 2009;34(4):437-40.	Outcome
214	Paternoster DM, Cardascia L, Narne S, Parise A, De Paoli M, Plebani M, et al. Predictive markers of preterm labour. Cosmi EV, editor1999. 463-7 p.	Outcome
215	Paternoster DM, Muresan D, Vitulo A, Serena A, Battagliarin G, Dell'Avanzo M, et al. Cervical phIGFBP-1 in the evaluation of the risk of preterm delivery. Acta Obstetricia Et Gynecologica Scandinavica. 2007;86(2):151-5.	Outcome
216	Rahkonen L, Unkila-Kallio L, Nuutila M, Sainio S, Saisto T, Rutanen EM, et al. Cervical length measurement and cervical phosphorylated insulin-like growth factor binding protein-1 testing in prediction of preterm birth in patients reporting uterine contractions. Acta Obstetricia Et Gynecologica Scandinavica. 2009;88(8):901-8.	Outcome
217	Vetr M, Kudela M, Prasilova J. Fetal fibronectin in patients at increased risk for premature birth. Acta Universitatis Palackianae Olomucensis Facultatis Medicae. 1996;140(0):55-7.	Outcome
218	Centra M, Coata G, Picchiassi E, Alfonsi L, Meniconi S, Bini V, et al. Evaluation of quantitative fFn test in predicting the risk of preterm birth. Journal of Perinatal Medicine. 2017;45(1):91-8.	Outcome
219	fFN testing not always so accurate. Contemporary OB/GYN. 2009;54(10):18-.	Population
220	Abbott D, Chandiramani M, Seed P, Tribe R, Shennan A. Quantification of Fetal Fibronectin Improves the Accuracy of Prediction of Spontaneous Preterm Birth in High Risk Asymptomatic Women. Reproductive Sciences. 2013;20(S3):124A-A.	Population
221	Abbott D, Hezelgrave N, Seed P, Bennett P, Chandiramani M, David A, et al. Equipp: Evaluation of fetal fibronectin with a novel bedside quantitative instrument for the prediction of preterm birth. Journal of Maternal-Fetal and Neonatal Medicine. 2014;27:368-9.	Population
222	Abbott D, Smout E, Foster C, Van der Westhuizen M, Seed P, Shennan AH. The development and validation of a prediction tool for spontaneous preterm birth incorporating fetal fibronectin and cervical length. Bjog-an International Journal of Obstetrics and Gynaecology. 2012;119(6):e1-e.	Population
223	Audibert F, Fortin S, Delvin E, Djemli A, Brunet S, Dube J, et al. Contingent Use of Fetal Fibronectin Testing and Cervical Length Measurement in Women With Preterm Labour. Journal of Obstetrics and Gynaecology Canada. 2010;32(4):307-12.	Population

No.	Citation	Reason for Exclusion
224	Calvert B, Hezelgrave N, Seed P, Shennan A. Quantitative fetal fibronectin and birth outcome in women with previous cervical surgery and a short cervix. Bjog-an International Journal of Obstetrics and Gynaecology. 2015;122:102-3.	Population
225	Esplin M. 6: The use of serial cervical length and quantitative fetal fibronectin to identify nulliparous women at risk of subsequent spontaneous preterm birth. American Journal of Obstetrics & Gynecology. 2016;214:S4-S5.	Population
226	Esplin MS, Elovitz MA, Iams JD, Parker CB, Wapner RJ, Grobman WA, et al. Predictive Accuracy of Serial Transvaginal Cervical Lengths and Quantitative Vaginal Fetal Fibronectin Levels for Spontaneous Preterm Birth Among Nulliparous Women. JAMA: Journal of the American Medical Association. 2017;317(10):1047-56.	Population
227	Goldenberg R, Iams J, Mercer M, Meis P, Mowad A, Copper R, et al. Fetal fibronectin and spontaneous preterm birth. American journal of obstetrics and gynecology. 1995;172(1 PART 2):254.	Population
228	Goldenberg RI. The preterm prediction study: Patterns of cervicovaginal fetal fibronectin (FFN) as predictors of subsequent FFN and spontaneous preterm delivery (SPD). American journal of obstetrics and gynecology. 1997;176(1 PART 2):S53.	Population
229	Goldenberg RL. The preterm prediction study: Sequential cervical length and fetal fibronectin testing for the prediction of spontaneous preterm birth (SPB). American journal of obstetrics and gynecology. 1999;180(1 PART 2):S27.	Population
230	Hughes K, Sim S, Roman A, Michalak K, Kane S, Sheehan P. Outcomes and predictive tests from a dedicated specialist clinic for women at high risk of preterm labour: A ten year audit. Australian and New Zealand Journal of Obstetrics and Gynaecology. 2017.	Population
231	Jeurgens-Borst A, Bekkers RLM, Sporken JMJ, van den Berg PP. Use of insulin like growth factor binding protein-1 in the diagnosis of ruptured fetal membranes. European Journal of Obstetrics Gynecology and Reproductive Biology. 2002;102(1):11-4.	Population
232	Kalafat E, Yuce T, Tanju O, Koc A. Preterm premature rupture of membrane assessment via transperineal ultrasonography: a diagnostic accuracy study. Journal of Maternal-Fetal & Neonatal Medicine. 2016;29(22):3690-4.	Population
233	Kwek K, Khi C, Ting HS, Yeo GSH. Evaluation of a bedside test for phosphorylated insulin-like growth factor binding protein-1 in preterm labour. Annals Academy of Medicine Singapore. 2004;33(6):780-3.	Population
234	Mesic ogic L, Micic D, Omeragic F, Kovac R, Fazlagic S. IGFBP-1 marker of cervical ripening and predictor of preterm birth. Medicinski Glasnik Ljekarske Komore Zenickodobojskog Kantona. 2016;13(2):118-24.	Population
235	Paternoster D, Pignatara R, Grella P. Insulin-like growth factor protein test (IGFBP-1) in predicting preterm delivery. Cosmi EV, editor1998. 375-8 p.	Population
236	Paternoster DM, Pignataro R, Stella A, Bertoldini M, Bracciante M. New biotechnologies in the prediction of preterm birth. Cosmi EV, editor2000. 457-64 p.	Population
237	The use of fetal fibronectin testing in clinical practice: an audit of referrals for threatened preterm labour to the Victorian perinatal emergency retrieval service from non-tertiary hospitals. Journal of Paediatrics & Child Health. 2017;53:90-.	Test
238	Abbott D, Radford S, Seed P, Shennan AH. The first report of a quantitative fetal fibronectin bedside analyser in the prediction of preterm birth in symptomatic women. Bjog-an International Journal of Obstetrics and Gynaecology. 2012;119(6):e1-e2.	Test
239	Abenhaim HA, Morin L, Benjamin A. Does availability of fetal firbonectin testing in the management of threatened preterm labour affect the utilization of hospital resources? Journal of Obstetrics & Gynaecology Canada: JOGC. 2005;27(7):689-94.	Test
240	Ai F, Li G-Q, Jiang J, Dong X-D. Neutrophil elastase and fetal fibronectin levels as predictors of single-birth prematurity. Experimental and Therapeutic Medicine. 2015;10(2):665-70.	Test
241	Akers A, Jarzembowski JA, Johnson CT, Lieberman RW, Dalton VK. Examining the relationship between positive mid-gestational fetal fibronectin assays and histological evidence of acute placental inflammation. Journal of Perinatal Medicine. 2007;35(1):36-42.	Test
242	Allen KA, Kiwi R. Clinical utilization of a rapid fetal fibronectin assay. Annals of Clinical and Laboratory Science. 2000;30(3):323-4.	Test
243	Andersen HF. Use of fetal fibronectin in women at risk for preterm delivery. Clinical Obstetrics and Gynecology. 2000;43(4):746-58.	Test

No.	Citation	Reason for Exclusion
244	Anwar A, Lindow SW, Greaves L, Hall S, Jha R. The use of fetal fibronectin in suspected pre-term labour. Journal of Obstetrics & Gynaecology. 2014;34(1):45-7.	Test
245	Armson A, Dodds L, Dooley K, Howlett A, McPhee A, Scott H. Fetal fibronectin testing for suspected preterm labour in Nova Scotia. American journal of obstetrics and gynecology. 2004;191(6):S115-S.	Test
246	Ayala-Mendez JA, Kurtzman J, Rosales-Ortiz S, Martinez-Alvarez O, Das A, Jimenez-Solis G. FETAL FIBRONECTIN STATUS MARKEDLY MODIFIES THE RISK OF PRETERM DELIVERY IN LOW RISK PATIENTS WITH SYMPTOMATIC PRETERM LABOR AND CERVICAL SHORTENING. American journal of obstetrics and gynecology. 2008;199(6):S235-S.	Test
247	Bare T, Manoku N. Fetal fibronectin and cericometry in patients at increased risk for premature birth. Acta Microbiologica Hellenica. 2015;60 (3):186.	Test
248	Bartnicki J, Casal D, Kreaden US, Saling E, Vetter K. Fetal fibronectin in vaginal specimens predicts preterm delivery and very-low-birth-weight infants. American journal of obstetrics and gynecology. 1996;174(3):971-4.	Test
249	Bayar E, Wietek N, Page L, Girling J. A service evaluation of quantitative fetal fibronectin (qfFN) in symptomatic women as a predictor of preterm birth at West Middlesex University Hospital (WMUH): potential to reduce bed occupancy further. Bjog-an International Journal of Obstetrics and Gynaecology. 2015;122:97-.	Test
250	Benson JE, Landy HJ, Ghidini A, Drassinower D, Poggi SH. Fetal fibronectin for evaluation of preterm labor in the setting of cervical cerclage. Journal of Maternal-Fetal & Neonatal Medicine. 2012;25(11):2330-2.	Test
251	Berghella V, Hayes E, Visintine J, Baxter J. Fetal fibronectin testing for reducing the risk of preterm birth. Cochrane Database of Systematic Reviews. 2007;(4) (no pagination)(CD006843).	Test
252	Berghella V, Hayes E, Visintine J, Baxter JK. Fetal fibronectin testing for reducing the risk of preterm birth. Cochrane Database of Systematic Reviews. 2008:N.PAG-N.PAG.	Test
253	Berghella V, Saccone G. Fetal fibronectin testing for prevention of preterm birth in singleton pregnancies with threatened preterm labor: a systematic review and metaanalysis of randomized controlled trials. American Journal of Obstetrics & Gynecology. 2016;215(4):431-8.	Test
254	Bolt L, Chandiramani M, De Greeff A, Shennan A. Does fetal fibronectin testing change patient management in women at risk of pre-term labour? International Journal of Gynecology and Obstetrics. 2009;107:S527-S8.	Test
255	Boots AB, Sanchez-Ramos L, Bowers DM, Kaunitz AM, Zamora J, Schlattmann P. The short-term prediction of preterm birth: a systematic review and diagnostic metaanalysis. American Journal of Obstetrics & Gynecology. 2014;210(1):54.e1-.e10.	Test
256	Bruijn M, Vis J, Wilms F, Oudijk M, Kwee A, Porath M, et al. The contribution of vaginal examination to risk stratification of women with signs of preterm labor before 34 weeks gestation: the APOSTEL1-cohort. American journal of obstetrics and gynecology. 2014;210(1, Suppl. S):S362-S3.	Test
257	Burrus D, Adair D, Aptel K, Ernest J, Frye A, Veille J. The effect of digital cervical exam on measured levels of fetal fibronectin. American journal of obstetrics and gynecology. 1996;174(1 PART 2):352.	Test
258	Burrus DR, Ernest JM, Veille JC. FETAL FIBRONECTIN, INTERLEUKIN-6, AND C-REACTIVE PROTEIN ARE USEFUL IN ESTABLISHING PROGNOSTIC SUBCATEGORIES OF IDIOPATHIC PRETERM LABOR. American journal of obstetrics and gynecology. 1995;173(4):1258-62.	Test
259	Burwick R, Zork N, Lee G, Ross M, Kjos S. Cervilenz assessment of cervical length compared to fetal fibronectin in the prediction of preterm delivery in women with threatened preterm labor. Journal of maternal-fetal & neonatal medicine [Internet]. 2011; 24(1):[127-31 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/260/CN-00791260/frame.html.	Test
260	Chandiramani M, Das A, Bolt LA, de Greeff A, Filmer J, Shennan A. Assessment of Fetal Fibronectin Testing to Improve Preterm Birth Management. Reproductive Sciences. 2014;21(3):151A-2A.	Test
261	Cheung KW, Ngu SF, Lee CP. Fetal fibronectin test on Chinese women with symptoms of preterm labour: a pilot study. Hong Kong Medical Journal. 2013;19(5):424-8.	Test

No.	Citation	Reason for Exclusion
262	Chien PF, Khan KS, Ogston S, Owen P. The diagnostic accuracy of cervico-vaginal fetal fibronectin in predicting preterm delivery: an overview (Structured abstract). British Journal of Obstetrics and Gynaecology [Internet]. 1997; 104(4):[436-44 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/cldare/articles/DARE-11997000650/frame.html.	Test
263	Clock C, Kurtzman J. Positive fetal fibronectin status is associated with dynamic cervical shortening in patients with normal baseline cervical length and symptomatic contractions. American journal of obstetrics and gynecology. 2007;197(6):S199-S.	Test
264	Coleman MAG, Keelan JA, McCowan LME, Townend KM, Mitchell MD. Predicting preterm delivery: comparison of cervicovaginal interleukin (IL)-1 beta, IL-6 and IL-8 with fetal fibronectin and cervical dilatation. European Journal of Obstetrics Gynecology and Reproductive Biology. 2001;95(2):154-8.	Test
265	Cox S, Little B, Dax J, Leveno K. Fetal fibronectin and preterm delivery. American journal of obstetrics and gynecology. 1996;174(1 PART 2):306.	Test
266	Defranco EA, Lewis DF, Odibo AO. Improving the screening accuracy for preterm labor: is the combination of fetal fibronectin and cervical length in symptomatic patients a useful predictor of preterm birth? A systematic review. American Journal of Obstetrics & Gynecology. 2013;208(3):233.e1-6.	Test
267	Deshpande S, van Asselt A, Tomini F, Armstrong N, Allen A, Noake C, et al. Rapid fetal fibronectin testing to predict preterm birth in women with symptoms of premature labour: a systematic review and cost analysis. Health Technology Assessment. 2013;17(32):1-138.	Test
268	Diaz J, Chedraui P, Hidalgo L, Medina M. The clinical utility of fetal fibronectin in the prediction of pre-term birth in a low socio-economic setting hospital in Ecuador. Journal of Maternal-Fetal & Neonatal Medicine. 2009;22(2):89-93.	Test
269	Dutta D, Norman J. Pilot study into the efficacy of foetal fibronectin testing in minimising hospital admissions in women presenting with symptoms of preterm labour: a randomised controlled trial of obstetric and neonatal outcomes. Archives of gynecology and obstetrics [Internet]. 2011; 284(3):[559-65 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/101/CN-00813101/frame.html.	Test
270	Ehsanipoor RM, Swank ML, Jwa SC, Wing DA, Tarabulsi G, Blakemore KJ. Placental alpha-Microglobulin-1 in Vaginal Secretions of Women with Evidence of Preterm Labor. American Journal of Perinatology. 2016;33(2):208-13.	Test
271	Elliot J, Maharajan P, Thampi P. Test that predicts premature labour: Efficacy of foetal fibronectin test and its cost implications. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2012;97:A19-A20.	Test
272	Farag AH, Mohammed MM, Ellaithy MI, Salama HA. Blind vaginal fetal fibronectin swab for prediction of preterm birth. Journal of Obstetrics & Gynaecology Research. 2015;41(7):1009-17.	Test
273	Fell DB, Sprague AE, Grimshaw JM, Yasseen IAS, Coyle D, Dunn SI, et al. Evaluation of the impact of fetal fibronectin test implementation on hospital admissions for preterm labour in Ontario: A multiple baseline time-series design. BJOG: An International Journal of Obstetrics and Gynaecology. 2014;121(4):438-46.	Test
274	Foong E, Chye Chuah S. Clinical application of quantitative fetal fibronectin testing in the management of preterm labour at a regional hospital. Australian and New Zealand Journal of Obstetrics and Gynaecology. 2016;56:37-8.	Test
275	Foxman EF, Jarolim P. Use of the fetal fibronectin test in decisions to admit to hospital for preterm labor. Clinical Chemistry. 2004;50(3):663-5.	Test
276	Franco R, Shaw K, Williams C, Hickok D. Risk of preterm delivery with a positive fetal fibronectin test result and a normal cervical length. Obstetrics and Gynecology. 2005;105(4):114S-S.	Test
277	Fricke E. A quality analysis of the evaluation of preterm labor. American journal of obstetrics and gynecology. 2015;1):S66.	Test
278	Gaikwad VA, Rajput UG, Puri MS. Cervico vaginal fluid fibronectin level in assessing risk of preterm labour. Research Journal of Pharmaceutical, Biological and Chemical Sciences. 2014;5(3):156-62.	Test
279	Gao L, Zhang JP, Chen H, Guo ZJ, Chen LB, Tan JP, et al. Fetal fibronectin detection for preterm birth prediction. Genetics and Molecular Research. 2014;13(1):1323-8.	Test

No.	Citation	Reason for Exclusion
280	Garite T, Lucah AC, Williams C, Bobritchi B, Lapointe J. A comparison of speculum and non-speculum collection of cervicovaginal SWAB specimens for fetal fibronectin testing. American journal of obstetrics and gynecology. 2005;193(6):S52-S.	Test
281	Giles W, Bisits A, Knox M, Madsen G, Smith R. The effect of fetal fibronectin testing on admissions to a tertiary maternal-fetal medicine unit and cost savings. American Journal of Obstetrics & Gynecology. 2000;182(2):439-42.	Test
282	Goepfert AR, Andrews WW. The preterm prediction study: Association between cervical interleukin-6 (IL-6), fetal fibronectin (FFN), and spontaneous preterm birth (SPTB). American journal of obstetrics and gynecology. 1997;176(1 PART 2):S6.	Test
283	Goffeng AR, Holst E, Milsom I, Lindstedt G, Lundberg P-A, Andersch B. Factors influencing vaginal fetal fibronectin concentrations during pregnancy. Gynecologic and Obstetric Investigation. 1999;47(2):83-8.	Test
284	Goffinet F, Kayem G, Maillard F, Trebeden H, Cabrol D, Weill B, et al. Detection of interleukin 6 mRNA by RT-PCR in vaginal secretions: association with preterm delivery and neonatal infection in women with preterm labour and intact membranes. European Journal of Obstetrics Gynecology and Reproductive Biology. 2005;123(2):167-73.	Test
285	Golic M, Siedentopf J-P, Pauly F, Hinkson L, Henrich W, Tucher E. Influence of transvaginal ultrasound examination on quantitative vaginal fibronectin measurements: a prospective evaluation study. Journal of Perinatal Medicine. 2017;45(1):85-9.	Test
286	Grobman W, Welshman E, Calhoun E. Does fetal fibronectin use in the diagnosis of preterm labor affect physician behavior and health care costs? A randomized trial. American journal of obstetrics and gynecology [Internet]. 2004; 191(1):[235-40 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/981/CN-00480981/frame.html.	Test
287	Groom KM, Liu E, Allenby K. The impact of fetal fibronectin testing for women with symptoms of preterm labour in routine clinical practice within a New Zealand population. Australian & New Zealand Journal of Obstetrics & Gynaecology. 2006;46(5):440-5.	Test
288	Harstall C. Using Fetal Fibronectin to Diagnose Pre-term Labour: The role of rapid fetal fibronectin assay in the management of spontaneous preterm labour (Structured abstract). Health Technology Assessment Database [Internet]. 2008; (4):[79 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32008100049/frame.html.	Test
289	Helmer H, Brunbauer M, Rohrmeister K. Exploring the role of Tractocile in everyday clinical practice. Bjog-an International Journal of Obstetrics and Gynaecology. 2003;110:113-5.	Test
290	Heng YJ, Pennell CE, Chua HN, Perkins JE, Lye SJ. Whole blood gene expression profile associated with spontaneous preterm birth in women with threatened. PLoS ONE. 2014;9 (5) (no pagination)(e96901).	Test
291	Hermans FJR, Bruijn MMC, Vis JY, Wilms FF, Oudijk MA, Porath MM, et al. Risk stratification with cervical length and fetal fibronectin in women with threatened preterm labor before 34weeks and not delivering within 7days. Acta Obstetricia Et Gynecologica Scandinavica. 2015;94(7):715-21.	Test
292	Hincz P, Wilczynski J, Kozarzewski M, Borowski D, Krekora M, Szaflik K. Two-step test: Fetal fibronectin and sonographic examination of the uterine cervix in patients presenting with preterm contractions. American journal of obstetrics and gynecology. 2000;182(1 Part 2):S44.	Test
293	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 Obstetricia Et Gynecologica Scandinavica. 2002;81(1):58-63.	Test
294	Honest H, Forbes C, Durée K, Norman G, Duffy S, Tsourapas A, et al. Screening to prevent spontaneous preterm birth: systematic reviews of accuracy and effectiveness literature with economic modelling. Health Technology Assessment. 2009;13(37):1-627.	Test
295	Iams JD, Casal D, McGregor JA, Goodwin TM, Kreaden US, Lowensohn R, et al. FETAL FIBRONECTIN IMPROVES THE ACCURACY OF DIAGNOSIS OF PRETERM LABOR. American journal of obstetrics and gynecology. 1995;173(1):141-5.	Test
296	Ibrahim MI, Sherif A, El-Kady M, Ellaithy M, Husseiny A, Kamal M, et al. Can three-dimensional ultrasound measurement of fetal adrenal gland enlargement predict preterm birth? Archives of Gynecology and Obstetrics. 2015;292(3):569-78.	Test
297	Incerti M, Ghidini A, Korker V, Pezzullo JC. Performance of cervicovaginal fetal fibronectin in a community hospital setting. Archives of Gynecology and Obstetrics. 2007;275(5):347-51.	Test

No.	Citation	Reason for Exclusion
298	Joffe G, Jacques D, Burton B, Bemis-Heys R, Shelburne P, Lovelace. Impact of fetal fibronectin assay on admissions for preterm labor. American journal of obstetrics and gynecology. 1998;178(1 PART 2):S11.	Test
299	Joffe GM, Jacques D, Bemis-Heys R, Burton R, Skram B, Shelburne P. Impact of the fetal fibronectin assay on admissions for preterm labor. American journal of obstetrics and gynecology. 1999;180(3):581-6.	Test
300	Karamisheva V, Ivanov S, Nachev A. COMPARISON OF TWO RAPID STRIP TESTS BASED ON FFN AND PAMG-1 FOR THE PREDICTION OF IMMINENT SPONTANEOUS PRETERM DELIVERY. Comptes Rendus De L Academie Bulgare Des Sciences. 2016;69(5):661-6.	Test
301	Keeler S, Roman A, Coletta J, Kiefer D, Rust O. Fetal fibronectin testing prior to ultrasound-indicated CERC1 age: Can it predict procedure success? American journal of obstetrics and gynecology. 2007;197(6):S199-S.	Test
302	Keeler SM, Roman AS, Coletta JM, Kiefer DG, Feuerman M, Rust OA, et al. Fetal fibronectin testing in patients with short cervix in the midtrimester: can it identify optimal candidates for ultrasound-indicated cerclage? American Journal of Obstetrics & Gynecology. 2009;200(2):158.e1-6.	Test
303	Langer B, Boudier E, Schlaeder G. Cervico-vaginal fetal fibronectin: Predictive value during false labor. Acta Obstetricia Et Gynecologica Scandinavica. 1997;76(3):218-21.	Test
304	Lee G, Burwick R, Zork N, Kjos S. Does the use of fetal fibronectin in an algorithm for preterm labor reduce triage evaluation times? The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians [Internet]. 2013; 26(7):[706-9 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/772/CN-00966772/frame.html.	Test
305	Lee SM, Romero R, Park JW, Kim SM, Park CW, Korzeniewski SJ, et al. The clinical significance of a positive Amnisure test in women with preterm labor and intact membranes. Journal of Maternal-Fetal & Neonatal Medicine. 2012;25(9):1690-8.	Test
306	Leitich H, Egarter C, Kaider A, Hoblagschwandtner M, Berghammer P, Husslein P. Cervicovaginal fetal fibronectin as a marker for preterm delivery: A meta-analysis. American journal of obstetrics and gynecology. 1999;180(5):1169-76.	Test
307	Lopez RL, Francis JA, Garite TJ, Dubyak JM. Fetal fibronectin detection as a predictor of preterm birth in actual clinical practice. American journal of obstetrics and gynecology. 2000;182(5):1103-6.	Test
308	Lowe MP, Zimmerman B, Hansen W. Prospective randomized controlled trial of fetal fibronectin on preterm labor management in a tertiary care center. American journal of obstetrics and gynecology. 2004;190(2):358-62.	Test
309	Lu GC, Goldenberg RL, Cliver SP, Kreaden US, Andrews WW. Vaginal fetal fibronectin levels and spontaneous preterm birth in symptomatic women. Obstetrics and Gynecology. 2001;97(2):225-8.	Test
310	MacDonald WA, Bender M, Saxton A. Use of fetal fibronectin in the management of preterm labour in Nunavut. Alaska medicine. 2006;49(2 Suppl):215-7.	Test
311	Malak TM, Sizmur F, Bell SC, Taylor DJ. Fetal fibronectin in cervicovaginal secretions as a predictor of preterm birth. British Journal of Obstetrics and Gynaecology. 1996;103(7):648-53.	Test
312	Marvin KW, Keelan JA, Coleman MAG, McCowan LME, Zhou RL, Mitchell MD. Intercellular adhesion molecue-1 (ICAM-1) in cervicovaginal fluid of women presenting with preterm labor: Predictive value for preterm delivery. American Journal of Reproductive Immunology. 2000;43(5):264-71.	Test
313	Mateus J, Pereira L, Baxter J, Berghella V, Tolosa J. Effectiveness of fetal fibronectin testing compared with digital cervical assessment of women with preterm contractions. American Journal of Perinatology. 2007;24(6):381-5.	Test
314	Morrison JC, Allbert JR, McLaughlin BN, Whitworth NS, Roberts WE, Martin RW. ONCOFETAL FIBRONECTIN IN PATIENTS WITH FALSE LABOR AS A PREDICTOR OF PRETERM DELIVERY. American journal of obstetrics and gynecology. 1993;168(2):538-42.	Test
315	Most O, Langer O, Kerner R, Ben David G, Calderon I. Can myometrial electrical activity identify patients in preterm labor? American journal of obstetrics and gynecology. 2008;199(4).	Test

No.	Citation	Reason for Exclusion
316	Musaad SM, Melson CL, Boswell DR. Assessment of the impact of introducing fetal fibronectin assay in the management of preterm labour at Middlemore Hospital, New Zealand. Pathology. 2005;37(3):226-30.	Test
317	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. American journal of obstetrics and gynecology. 2007;197(4).	Test
318	Ness A, Visintine J, Ricci E, Boyle K, Berghella V. Use of fetal fibronectin and transvaginal ultrasound cervical length to triage women with suspected preterm labor: a randomized trial. American journal of obstetrics and gynecology [Internet]. 2006; 195(6 Suppl 1):[S67 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/777/CN-00582777/frame.html.	Test
319	Nguyen T. The cost-effectiveness of fetal fibronectin testing in suspected preterm labor: a randomized trial. Obstetrics and gynecology [Internet]. 2002; 99(4 Suppl):[97s p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/633/CN-00387633/frame.html.	Test
320	Owen P, Scott A. Can fetal fibronectin testing improve the management of preterm labour? Clinical and Experimental Obstetrics and Gynecology. 1997;24(1):19-22.	Test
321	Parisaei M, Currie J, O'Gorman N, Morris S, David AL. Implementation of foetal fibronectin testing: Admissions, maternal interventions and costs at 1year. Journal of Obstetrics and Gynaecology. 2016;36(7):888-92.	Test
322	Parker J, Bell R, Brennecke S. Fetal fibronectin in the cervicovaginal fluid of women with threatened preterm labour as a predictor of delivery before 34 weeks' gestation. Australian and New Zealand Journal of Obstetrics and Gynaecology. 1995;35(3):257-61.	Test
323	Parry E, Singh TG, Dow D, Noova F. Improved management in threatened preterm labour with rapid fetal fibronectin testing. Australian and New Zealand Journal of Obstetrics and Gynaecology. 2006;46(3):240-1.	Test
324	Plaut M, Smith W, Kennedy K. Fetal fibronectin: the impact of a rapid test on the treatment of women with preterm labor symptoms. American journal of obstetrics and gynecology [Internet]. 2003; 188(6):[1588-93; discussion 93-5 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/668/CN-00438668/frame.html.	Test
325	Plaut MM, Smith W, Kennedy K, Nageotte M, DeCastro E, Steinke R, et al. Fetal fibronectin: The impact of a rapid test on the treatment of women with preterm labor symptoms. American journal of obstetrics and gynecology. 2003;188(6):1588-95.	Test
326	Revah A, Hannah ME, Sue-A-Quan AK. Fetal fibronectin as a predictor of preterm birth: An overview. American Journal of Perinatology. 1998;15(11):613-21.	Test
327	Rinehart BK, Terrone DA, Isler CM, Barrilleaux PS, Bufkin L, Morrison JC. Pregnancy outcome in women with preterm labor symptoms without cervical change. American Journal of Obstetrics & Gynecology. 2001;184(5):1004-7.	Test
328	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. American journal of obstetrics and gynecology. 1996;175(5):1146-51.	Test
329	Roman AS, Koklanaris N, Paidas MJ, Mulholland J, Levitz M, Rebarber A. 'Blind' vaginal fetal fibronectin as a predictor of spontaneous preterm delivery. Obstetrics & Gynecology. 2005;105(2):285-9.	Test
330	Roos C, Schuit E, Scheepers HCJ, Bloemenkamp KWM, Bolte AC, Duvekot HJJ, et al. Predictive Factors for Delivery within 7 Days after Successful 48-Hour Treatment of Threatened Preterm Labor. Ajp Reports. 2015;5(2):E141-E9.	Test
331	Roos C, Vis JY, Scheepers HCJ, Bloemenkamp KWM, Duvekot HJJ, van Eyck J, et al. Fetal fibronectin status and cervical length in women with threatened preterm labor and the effectiveness of maintenance tocolysis. Journal of Maternal-Fetal & Neonatal Medicine. 2016;29(10):1556-61.	Test
332	Rose CH, McWeeney DT, Brost BC, Davies NP, Watson WJ. Cost-effective standardization of preterm labor evaluation. American journal of obstetrics and gynecology. 2010;203(3).	Test
333	Rossel Goffeng A, Holst E, Milsom I, Lindstedt G, Lundberg PA, Andersch B. Factors influencing vaginal fetal fibronectin concentrations during pregnancy. Gynecologic & Obstetric Investigation. 1999;47(2):83-8.	Test

No.	Citation	Reason for Exclusion
334	Rozenberg P, Goffinet F, Malagrida L, Giudicelli Y, Perdu M, Houssin I, et al. Evaluating the risk of preterm delivery: a comparison of fetal fibronectin and transvaginal ultrasonographic measurement of cervical length. American journal of obstetrics and gynecology. 1997;176(1 Pt 1):196-9.	Test
335	Sakai M, Sasaki Y, Yamagishi N, Tanebe K, Yoneda S, Saito S. The preterm labor index and fetal fibronectin for prediction of preterm delivery with intact membranes. Obstetrics & Gynecology. 2003;101(1):123-8.	Test
336	Salcedo A, Kurtzman J. Digital cervical exam characteristics stratify the risk of preterm birth based on fetal fibronectin status. American journal of obstetrics and gynecology. 2013;208(1):S203-S.	Test
337	Senden IPM, Owen P. Comparison of cervical assessment, fetal fibronectin and fetal breathing in the diagnosis of preterm labour. Clinical and Experimental Obstetrics and Gynecology. 1996;23(1):5-9.	Test
338	Skoll A, St Louis P, Amiri N, Delisle MF, Lalji S. The Evaluation of the Fetal Fibronectin Test for Prediction of Preterm Delivery in Symptomatic Patients. Journal of Obstetrics and Gynaecology Canada. 2006;28(3):206-13.	Test
339	Stevens AO, Chauhan SP, Magann EF, Martin RW, Bofill JA, Cushman JL, et al. Fetal fibronectin and bacterial vaginosis are associated with preterm birth in women who are symptomatic for preterm labor. American journal of obstetrics and gynecology. 2004;190(6):1582-9.	Test
340	Stuart RW, Jones RO, Underwood F, Kryka R, Bodor GS. The utility of fetal fibronectin in preterm labor decision making. American Journal of Clinical Pathology. 2007;128(3):513-4.	Test
341	Sukchaya K, Phupong V. A comparative study of positive rate of placental alpha-microglobulin-1 test in pre-term pregnant women with and without uterine contraction. Journal of Obstetrics & Gynaecology. 2013;33(6):566-8.	Test
342	Sullivan A, Hueppchen NA, Satin AJ. Cost effectiveness of bedside fetal fibronectin testing varies according to treatment algorithm. Journal of Maternal-Fetal Medicine. 2001;10(6):380-4.	Test
343	Sunagawa S, Takagi K, Ono K, Miyachi K, Kikuchi A. Comparison of biochemical markers and cervical length for predicting preterm delivery. Journal of Obstetrics and Gynaecology Research. 2008;34(5):812-9.	Test
344	Swamy GK, Simhan HN, Gammill HS, Heine RP. Clinical utility of fetal fibronectin for predicting preterm birth. Journal of Reproductive Medicine. 2005;50(11):851-6.	Test
345	Tanir HM, Sener T, Yildiz Z. Cervicovaginal fetal fibronectin (FFN) for prediction of preterm delivery in symptomatic cases: a prospective study. Clinical and Experimental Obstetrics & Gynecology. 2008;35(1):61-4.	Test
346	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 in Obstetrics & Gynecology. 2005;26(7):699-706.	Test
347	Tekesin I, Marek S, Hellmeyer L, Reitz D, Schmidt S. Assessment of rapid fetal fibronectin in predicting preterm delivery. Obstetrics & Gynecology. 2005;105(2):280-4.	Test
348	Tekesin I, Wallwiener D, Schmidt S. The value of quantitative ultrasound tissue characterization of the cervix and rapid fetal fibronectin in predicting preterm delivery. Journal of Perinatal Medicine. 2005;33(5):383-91.	Test
349	Triana H, Saldivar D, Contreras N, Guzman A, Soria A, Farran I, et al. Fetal fibronectin in the diagnosis of preterm Labor. Cabero L, Carrera JM, editors2001. 599-602 p.	Test
350	Tsoi E, Akmal S, Geerts L, Jeffery B, Nicolaides KH. Sonographic measurement of cervical length and fetal fibronectin testing in threatened preterm labor. Ultrasound in Obstetrics & Gynecology. 2006;27(4):368-72.	Test
351	Tsoi E, Akmal S, Geerts L, Jeffery B, Nicolaides KH. Sonographic measurement of cervical length ad fetal fibronectin testing in threatened preterm labor. Obstetrical and Gynecological Survey. 2006;61(7):441-2.	Test
352	Tsourapas A, Roberts T, Barton P, Honest H, Forbes C, Hyde C, et al. An economic evaluation of alternative test-intervention strategies to prevent spontaneous pre-term birth in singleton pregnancies (Structured abstract). Acta Obstetricia et Gynecologica Scandinavica [Internet]. 2009; 88(12):[1319-30 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/cleed/articles/NHSEED-22010000199/frame.html.	Test

No.	Citation	Reason for Exclusion
353	van Baaren GJ, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM, et al. Cost-effectiveness of diagnostic testing strategies including cervical length measurement and fibronectin testing in women with symptoms of preterm labor. Ultrasound in Obstetrics & Gynecology. 2017:31.	Test
354	van Baaren G-J, Vis J, Wilms F, Oudijk M, Kwee A, Porath M, et al. The risk of preterm delivery in women with signs of preterm labor before 34 weeks who do not deliver within 7 days: the APOSTEL-1 cohort. American journal of obstetrics and gynecology. 2014;210(1, Suppl. S):S375.	Test
355	Volumenie JL, Guibourdenche J, Doridot V, Sibony O, Oury JF, Blot P, et al. Failure of cervical fibronectin to predict premature delivery in a population of monofetal pregnancies with idiopathic preterm labor. European Journal of Obstetrics Gynecology and Reproductive Biology. 2001;97(1):35-9.	Test
356	von Schoening D, Fischer T, von Tucher E, Slowinski T, Weichert A, Henrich W, et al. Cervical sonoelastography for improving prediction of preterm birth compared with cervical length measurement and fetal fibronectin test. Journal of Perinatal Medicine. 2015;43(5):531-6.	Test
357	Watson DL, Kim SJ, Humphrey MD. Study of cervicovaginal fetal fibronectin status to guide treatment of threatened preterm labour. Australian & New Zealand Journal of Obstetrics & Gynaecology. 1998;38(2):185-7.	Test
358	Watson HA, Carter J, Seed PT, Tribe RM, Shennan AH. The QUIPP app: a safe alternative to a treat-all strategy for threatened preterm labour. Ultrasound in Obstetrics & Gynecology. 2017:24.	Test
359	Wilms FF, van Baaren GJ, Vis JY, Oudijk MA, Kwee A, Porath MM, et al. Prescribing patterns of antenatal corticosteroids in women with threatened preterm labor. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2015;192:47-53.	Test
360	Wilms FF, Vis JY, de Wit-Zuurendonk L, Porath MM, Mol BWJ. What Is the Risk of Preterm Delivery after Arrested Preterm Labor? American Journal of Perinatology. 2015;32(1):63-9.	Test
361	Woodworth A, Moore J, G'Sell C, Verdoes A, Snyder JA, Morris L, et al. Diagnostic accuracy of cervicovaginal interleukin-6 and interleukin-6: Albumin ratio as markers of preterm delivery. Clinical Chemistry. 2007;53(8):1534-40.	Test
362	Yoneda S, Sakai M, Sasaki Y, Shiozaki A, Hidaka T, Saito S. Interleukin-8 and glucose in amniotic fluid, fetal fibronectin in vaginal secretions and preterm labor index based on clinical variables are optimal predictive markers for preterm delivery in patients with intact membranes. Journal of Obstetrics & Gynaecology Research. 2007;33(1):38-44.	Test
363	Yoneda S, Shiozaki A, Yoneda N, Shima T, Ito M, Yamanaka M, et al. Prediction of exact delivery time in patients with preterm labor and intact membranes at admission by amniotic fluid interleukin-8 level and preterm labor index. Journal of Obstetrics and Gynaecology Research. 2011;37(7):861-6.	Test
364	Yoneyama K, Kimura A, Kogo M, Kiuchi Y, Morimoto T, Okai T. Clinical predictive factors for preterm birth in women with threatened preterm labour or preterm premature ruptured membranes? Australian & New Zealand Journal of Obstetrics & Gynaecology. 2009;49(1):16-21.	Test
365	Magro-Malosso E, Seravalli V, Cozzolino M, Spitaleri M, Susini T, Di Tommaso M. Prediction of preterm delivery by fetal fibronectin in symptomatic and asymptomatic women with cervical length <= 20mm. Journal of Maternal-Fetal & Neonatal Medicine. 2017;30(3):294-7.	Test
366	Management of preterm labour. Volume 1: Evidence report and appendixes; Volume 2: Evidence tables (Structured abstract). Health Technology Assessment Database [Internet]. 2000; (4):[Vol 1: 224 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32001000946/frame.html.	Unobtainable
367	Fetal fibronectin for the prediction of preterm labor (Structured abstract). Health Technology Assessment Database. 2000(4).	Unobtainable
368	Fetal fibronectin point of care test (Project record). Health Technology Assessment Database [Internet]. 2006; (2). Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32006000075/frame.html.	Unobtainable
369	Actim partus test (Project record). Health Technology Assessment Database [Internet]. 2010; (4). Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32010001243/frame.html.	Unobtainable
370	Aboulenien W, Azzam A, Saleh F, Abollo M, Soliman A. Insulin-like growth factor binding protein-1 in cervico-vaginal secretions as an indicator of premature rupture of membranes: Comparison with nitrazine test and amniotic fluid index. International Journal of Gynecology and Obstetrics. 2009;107:S413.	Unobtainable

No.	Citation	Reason for Exclusion
371	American College Of O, Gynecologists. Fetal fibronectin preterm labor risk test. International Journal of Gynecology and Obstetrics. 1997;59(2):164.	Unobtainable
372	Anonymous. Fetal fibronectin measurement to predict preterm delivery. American Family Physician. 1992;45(1):287.	Unobtainable
373	Bare T, Mitre A, Vata E, Shkurti E, Hoxhaj O, Manoku N. BIOMARKERS AND CERVICAL LENGTH TO PREDICT SPONTANEOUS PRETERM BIRTH IN HIGH-RISK WOMEN. International Journal of Ecosystems and Ecology Science-IJEES. 2015;5(1):29-34.	Unobtainable
374	Basak S, Babbur V. Audit of in-utero transfers following introduction of rapid fetal fibronectin test for assessment of preterm labour in a district general hospital in United Kingdom. International Journal of Gynecology and Obstetrics. 2012;119:S291.	Unobtainable
375	Bertani D, Beski L, Tagliavini M, Benassi L. PROGNOSTIC ROLE OF FETAL FIBRONECTIN IN THREATENED PRETERM DELIVERY. Merialdi A, Cavatorta E, Gramellini D, editors1994. 193-6 p.	Unobtainable
376	Bianco M, Depew C, Meisel R. A fetal fibronectin assay for detection of rupture of membranes. American journal of obstetrics and gynecology. 1993;168(1 PART 2):435.	Unobtainable
377	Burwick R, York N, Lee, Ross M, Kjos S. Cervilenz assessment of cervical length is equivalent to fetal fibronectin in the prediction of preterm delivery. American college of obstetricians and gynecologists' annual meeting; 2010 may 17-19; san francisco, california, USA [Internet]. 2010. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/741/CN-00755741/frame.html.	Unobtainable
378	Crane J, Armson BA, Dodds L, Kennedy W, Feinberg R. Predictors of preterm delivery. American journal of obstetrics and gynecology. 1998;178(1 PART 2):S189.	Unobtainable
379	Gervasi MT, Marsoni V, Reveane A, Serra MM, Bracalente G, Darin G. Role of fetal fibronectin in the diagnosis of preterm delivery. Cosmi EV, editor1997. 23-8 p.	Unobtainable
380	Ghidini A, Poggi SH, Korker V. Perfomance of vaginal fetal fibronectin as predictor of preterm delivery in a community hospital. Journal of the Society for Gynecologic Investigation. 2006;13(2):338A-A.	Unobtainable
381	Gottschalk EM, Wenzel S, Salomon NS, Dudenhausen JW, Henrich W. Importance of cervical length measurement and fetal Fibronectin (fFN) to the prediction of lower premature birth rate. Geburtshilfe Und Frauenheilkunde. 2008;68:S25-S.	Unobtainable
382	Hayes, Inc. Fetal fibronectin test in women with symptoms of preterm labor (Structured abstract). Health Technology Assessment Database [Internet]. 2006; (4). Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32006001571/frame.html.	Unobtainable
383	Klee GG, Arnold SE, Heise RH. Fetal fibronectin as a risk predictor of preterm delivery. Clinical Chemistry. 2000;46(6):A65-A.	Unobtainable
384	Lu GC, Goldenberg RL, Cliver SP, Kreaden US, Andrews WW. Quantitative fetal fibronectin levels and the prediction of spontaneous preterm birth in symptomatic women. American journal of obstetrics and gynecology. 2000;182(1 Part 2):S53.	Unobtainable
385	Luzzi VI, Hankins K, Gronowski AM. Predictive value of fetal fibronectin as a biochemical marker of preterm delivery using the rapid fFN TLiTM system. Clinical Chemistry. 2002;48(6):A121-A.	Unobtainable
386	Macones GA, Stamilio D. Fetal fibronectin testing: The cost for false negative/positive results across gestational ages. American journal of obstetrics and gynecology. 2000;182(1 Part 2):S46.	Unobtainable
387	Marzolf G, Kreaden U, Lockhart T, Cooper M, Paulson M, Sheldon L, et al. Effect of time delay between specimen collection and testing with rapid fetal fibronectin (fFN) for the Adeza TLi (TM) System. Clinical Chemistry. 1999;45(6):A94-A.	Unobtainable
388	Mims O, Taylor S, Sundaram S, Stevens M, Manimekalai S, Goldstein PJ. Fetal fibronectin assay to predict pre term labor. Clinical Chemistry. 1999;45(6):A72-A.	Unobtainable
389	Min DL, Haberman S, McCalla S. The impact of the rapid fetal fibronectin assay on the management of preterm labor. Obstetrics and Gynecology. 2004;103(4):73S-S.	Unobtainable
390	Morgan BR. Clinical performance of the Adeza biomedical TLi Rapid Fetal Fibronectin Test at a multispecialty clinic. Clinical Chemistry. 2001;47(6):A145-A.	Unobtainable

No.	Citation	Reason for Exclusion
391	Nishida N, Asakura H, Yonezawa M, Tateyama N, Doi D, Fukami T. Prediction of preterm birth in comparison of TVS findings to those with positive fetal fibronectin (fFN). International Journal of Gynecology and Obstetrics. 2009;107:S468.	Unobtainable
392	Oyelese Y, Peters S, Ananth CV, Kinzler W, Smulian JC. The utility of fetal fibronectin in predicting preterm birth in clinical practice. Journal of the Society for Gynecologic Investigation. 2006;13(2):123A-A.	Unobtainable
393	Paternoster DM, Lazzarin L, Stella A, Pignataro R. Fetal fibronectin and determination of blood ferritin levels in preterm delivery. Cosmi EV, editor1997. 103-6 p.	Unobtainable
394	Pates JA, Lattu AL, Nielsen PE. The value of fetal fibronectin in predicting the onset of labor before 41 weeks of gestation. Obstetrics and Gynecology. 2008;111(4):102S-S.	Unobtainable
395	Paulson MC, Lapointe J, Sheldon E, Mason R, Hernandez O. Elimination of sample warming for rapid fFN testing. Clinical Chemistry. 2002;48(6):A194-A.	Unobtainable
396	Paulson MC, Sheldon EL. A Comparison of rapid fFN for the TLi system and the rapid fFN for the TLi IQ system. Clinical Chemistry. 2001;47(6):A111-A.	Unobtainable
397	Quintieri F, Paternoster D, Depaoli M, Grella PV, Plebani M. FETAL FIBRONECTIN IN EARLY DIAGNOSIS OF PRETERM BIRTHS. Merialdi A, Cavatorta E, Gramellini D, editors1994. 187-91 p.	Unobtainable
398	Ray D, Dyson D, Hendershott C, Field R, Walton D, Newman L, et al. A comparison of rapid fFN on the TLiTM system to the fFN enzyme immunoassay in symptomatic women. American journal of obstetrics and gynecology. 2000;182(1 Part 2):S44.	Unobtainable
399	Reveane A, Marsoni V, Gervasi MT, Amici G, Serra M, Marzolini M. Correlation between fetal fibronectin levels and the risk of premature delivery. Cosmi EV, editor1997. 85-8 p.	Unobtainable
400	Rinehart BK, Terrone DA, Isler CM, Barrilleaux PS, Bufkin LK, Morrison JC. Patients with false labor symptoms and positive fetal fibronectin are more likely to deliver preterm. American journal of obstetrics and gynecology. 2000;182(1 Part 2):S49.	Unobtainable
401	Sheldon EL, Paulson MC, Hernandez O. The effect of yeast on rapid fFN for the TLi(tm) system. Clinical Chemistry. 2001;47(6):A184-A.	Unobtainable
402	Sheldon EL, Paulson MC, Lockhart T, Marzolf G. Analytical evaluation of Rapid Fetal Fibronectin (fFN) for the Adeza Tli((TM)) System. Clinical Chemistry. 2000;46(6):A4-A.	Unobtainable
403	Siddiqa M, Haloob R. An audit of fetal fibronectin test. International Journal of Gynecology and Obstetrics. 2012;119:S752.	Unobtainable
404	Volpogni C, Facchinetti F, Martinez F, Genazzani AR. FETAL FIBRONECTIN AS PREDICTOR OF PRETERM BIRTH. Merialdi A, Cavatorta E, Gramellini D, editors1994. 183-6 p.	Unobtainable

A2.1 Raw rest accuracy data as reported in the papers

Table 55 DTA values for prediction of delivery within 48 hours

	N	True Positive	False Positive	True Negative	False Negative	Sensitivity n/N (%)	Specificity n/N (%)	PPV n/N (%)	NPV n/N (%)	+LR (95% CI)	-LR (95% CI)
Actim Partus											
Brik, 2010	276	NR	NR	NR	NR	(73.7)	(64.9)	(16.1)	(96.4)	2.10 (1.52, 2.91)	0.41 (0.19, 0.87)
Goyal, 2016	60	23	11	14	12	NR	NR	NR	NR	NR	NR
Lembet, 2002	36	14	4	17	1	(93.3)	(81)	(77.8)	(94.4)	4.9 (2.0, 11.9)	0.08 (0.01, 0.50)
Ting, 2007	94	NR	NR	NR	NR	(100)	(74)	(18)	(100)	NR	NR
Tripathi, 2016	468	NR	NR	NR	NR	(95.4)	(82.2)	(61.7)	(98.3)	NR	NR
Vishwekar 2017[a]	30	11	3	14	2	(73.3)	(64.3)	(68.8)	(69.2)	NR	NR
PartoSure											
Werlen, 2015	41	0	1	39	1	NR	NR	NR	NR	NR	NR

a, one patient with a negative result absconded so could not obtain delivery details

Table 56 DTA values for prediction of delivery within 7 days

		n	True Positive	False Positive	True Negative	False Negative	Sensitivity n/N (%)	Specificity n/N (%)	PPV n/N (%)	NPV n/N (%)	+LR (95% CI)	-LR (95% CI)
Bruijn, 2016: APOSTEL-1	fFN <10 ng/ml	350	66	162	119	3	NR	NR	NR	NR	NR	NR
	fFN <200 ng/ml		49	46	235	20	NR	NR	NR	NR	NR	NR
	fFN <500 ng/ml		29	12	269	40	NR	NR	NR	NR	NR	NR
	Actim Parus		54	30	251	15	(78.3)	(89.3)	(64.3)	(94.4)	NR	NR
Hadzi- Lega, 2017	Actim Partus	57	5	12	39	1	5/6 (83) [35.88, 99.58]	39/51 (76) [62.51, 87.21]	5/17 (29) [10.31, 55.96]	39/40 (98) [86.84, 99.94]	NR	NR
	PartoSure		5	5	46	1	5/6 (83) [35.88, 99.58]	46/51 (90) [78.59, 96.74]	5/10 (50) [18.71, 81.29]	46/47 (98) [88.71, 99.95]	NR	NR
Abo El-Ezz, 2014	Actim Partus	57	20	9	18	10	NR	NR	NR	NR	NR	NR
Altinkaya, 2009		105	9	16	75	5	NR	NR	NR	NR	NR	NR
Azlin, 2010		51	4	3	43	1	(80.0)	(93.5)	(57.1)	(97.7)	NR	NR

		n	True Positive	False Positive	True Negative	False Negative	Sensitivity n/N (%)	Specificity n/N (%)	PPV n/N (%)	NPV n/N (%)	+LR (95% CI)	-LR (95% CI)
Brik, 2010		276	NR	NR	NR	NR	73.1	66.2	21.8	95	2.16 [1.60, 2.92]	0.41 [0.21, 0.78]
Cooper, 2012		349	2	89	254	4	2/6 (33) [0.00, 0.71]	254/343 (74) [0.69, 0.79]	2/91 (2) [0.00, 0.05]	254/258 (98) [0.97, 1.00]	1.28 [0.41, 4.04]	0.90 [0.51, 1.59]
Danti, 2011		60	2	17	39	2	(50) [7, 93]	(70) [56, 81]	(11) [1, 33]	(95) [83, 99]	1.65 [0.57, 4.74]	0.72 [0.27, 1.94]
Eroglu, 2007		51	5	7	38	1	(83.3)	(84.4)	(41.7)	(97.4)	5.36 []2.3, 12.2]	0.20 [0.01, 0.7]
Goyal, 2016		60	26	8	8	18	NR	NR	NR	NR	NR	NR
Lembet, 2002		36	15	3	17	1	(93.8)	(85)	(83.3)	(94.1)	6.2 [2.2, 17.8]	0.07 [0.01, 0.5]
Riboni, 2011		210	NR	NR	NR	NR	(50)	(83.7)	(10.8)	(97.7)	NR	NR
Tanir, 2009[a]		68	14	11	42	1	14/15 (93.3)	42/53 (79.2)	14/25 (56)	42/43 (97.6)	4.4 [2.1, 5.2]	0.8 [0.4, 0.9]
Ting, 2007		94	NR	NR	NR	NR	(69)	(78)	(39)	(92)	NR	NR
Tripathi, 2016		468	NR	NR	NR	NR	(94.7)	(92.4)	(85.6)	(97.3)	NR	NR
Vishwekar 2017[b]		30	13	1	10	5	(68.4)	(90)	(92.9)	(60)	NR	NR
Bolotskikh, 2017	PartoSure	99	12	4	83	0	12/12 (100) [74, 100]	83/87 (95) [89, 99]	12/16 (75) [48, 93]	83/83 (100) [96, 100]	NR	NR
Nikolova, 2015		203	28	9	159	7	28/ 35 (80) [63.1, 91.6]	159/168 (95) [90.1, 97.5]	28/37 (76) [58.8, 88.2]	159/166 (96) [91.5, 98.3]	NR	NR
Werlen, 2015		41	0	1	39	1	(0) [0.0, 9.75]	(97.5) [86.8, 99.9]	(0) [0.0, 97.5]	(97.5) [86.8, 99.9]	NR	NR
Bruijn, 2016: EUIFS	fFN <10 ng/ml	455	45	276	131	3	NR	NR	NR	NR	NR	NR
	fFN <200 ng/ml	455	34	87	320	14	NR	NR	NR	NR	NR	NR
	fFN <500 ng/ml	455	14	23	384	34	NR	NR	NR	NR	NR	NR

a, 'when there were no visible lines, which was observed in two cases, a new sample was not taken. These patients were assigned test positive.' b, one patient with a negative result absconded so could not obtain delivery details

Appendix 3. Additional diagnostic test accuracy data on cervical length

In this overview, diagnostic test accuracy data, against a reference standard of preterm birth within 7 days, are provided for 1) cervical length ideally at the threshold of <15mm (using a threshold of <15mm is included in current NICE guidance to detect preterm labour). Where studies present cervical length data at various thresholds, only the data for the <15mm threshold are reported. If data were not available for the <15mm threshold, data for the closest reported threshold are reported. These data are presented in Table 60, together with test accuracy data for index tests from the same studies (i.e. data from index tests produced for the systematic review of PartoSure, Actim Partus, and quantitative fFN at thresholds other than 50ng/ml, see section 2.2).

Table 57 Test accuracy results (against preterm birth within 7 days) for index tests, cervical length measurement and fFN at 50ng/ml

Study	Test	Participants (*n*)	Sensitivity %; 95%CI	Specificity %; 95%CI	PPV %; 95%CI	NPV %; 95%CI
Azlin 2010 [50]	CL<25mm	51	80.0; 28.4-99.5	71.7; 56.5-84.0	23.5; 6.8-49.9	97.1;84.7-99.9
	Actim Partus	51	80.0; 28.4-99.5	93.5; 82.1-98.6	57.1; 18.4-90.1	97.7; 88.0-99.9
Bolotskikh, 2017 [61]	CL<15mm[a]	99	33.3; 9.9-65.1	98.9; 93.8-100.0	80.0; 28.4-99.5	91.5; 83.9-96.3
	PartoSure	99	100.0; 73.5-100.0	95.4; 88.6-98.7	75.0; 47.6-92.7	100.0; 95.7-100.0
APOSTEL-1 [45, 46]	CL<15mm[a]	350	72.5; 60.4-82.5	83.3; 78.4-87.4	51.5; 41.2-61.8	92.5; 88.5-95.4
	fFN@10	350	95.7; 87.8-99.1	42.3; 36.5-48.4	28.9; 23.2-35.3	97.5; 93.0-99.5
	fFN@50[c]	350	91.3; 82.0-96.7	64.8; 58.9-70.3	38.9; 31.3-46.9	96.8; 93.2-98.8
	fFN@200	350	71.0; 58.8-81.3	83.6; 78.8-87.8	51.6; 41.1-62.0	92.2; 88.1-95.1
	fFN@500	350	42.0; 30.2-54.5	95.7; 92.7-97.8	70.7; 54.5-83.9	87.1; 82.8-90.6
	Actim Partus	350	78.3; 66.7-87.3	89.3; 85.1-92.7	64.3;53.1-74.4	94.4; 90.9-96.8
EUIFS [64]	CL<15mm[a]	450	51.3; 34.8-67.6	81.8; 77.7-85.4	21.3; 13.5-30.9	94.6; 91.7-96.7
	fFN@10	455	93.8; 82.8-98.7	32.2; 27.7-37.0	14.0; 10.4-18.3	97.8; 93.6-99.5
	fFN@50[c]	455	89.6; 77.3-96.5	62.2; 57.3-66.9	21.8; 16.3-28.3	98.1; 95.5-99.4
	fFN@200	455	70.8; 55.9-83.0	78.6; 74.3-82.5	28.1; 20.3-37.0	95.8; 93.1-97.7
	fFN@500	455	29.2; 17.0-44.1	94.3; 91.6-96.4	37.8; 22.5-55.2	91.9; 88.8-94.3
Cooper, 2012 [1]	fFN@50[f]	291	33.3; 4.3-77.7	89.8; 85.7-93.1	6.5; 0.8-21.4	98.5; 96.1-99.6
	Actim Partus	349	33.3; 4.3-77.7	74.1; 69.1-78.6	2.2; 0.3-7.7	98.4; 96.1-99.6
Danti, 2011 [52]	CL <20mm[a] (sample 1)	60	75.0; 19.4-99.4	71.4; 57.8-82.7	15.8;3.4-39.6	97.6; 87.1-99.9
	CL <20mm[a] (sample 2)	102	75.0; 19.4-99.4	83.7; 74.8-90.4	15.8; 3.4-39.6	98.8; 93.5-100.0

	Actim Partus	60	50.0; 6.8-93.2	69.6; 55.9-81.2	10.5; 1.3-33.1	95.1;83.5-99.4
Eroglu, 2007 [53]	fFN@50[d]	51	83.3; 35.9-99.6	80.0; 65.4-90.4	35.7; 12.8-64.9	97.3; 85.8-99.9
	Actim Partus	51	83.3; 35.9-99.6	84.4; 70.5-93.5	41.7; 15.2-72.3	97.4; 86.5-99.9
	CL <20mm	51	66.7; 22.3-95.7	95.6; 84.9-99.5	66.7; 22.3-95.7	95.6; 84.9-99.5
Goyal, 2016 [54]	CL<25mm[b]	60	80.5; 65.1-91.2	31.6; 12.6-56.6	71.7; 56.5-84.0	42.9; 17.7-71.1
	Actim Partus	60	59.1; 43.2-73.7	50.0; 24.7-75.3	76.5; 58.8-89.3	30.8; 14.3-51.8
Hadzi-Lega, 2017 [47]	CL<25mm	57	100.0; 54.1-100.0	70.6; 56.2-82.5	28.6; 11.3-52.2	100.0; 90.3-100.0
	Actim Partus	57	83.3; 35.9-99.6	76.5; 62.5-87.2	29.4; 10.3-56.0	97.5; 86.8-99.9
	PartoSure	57	83.3; 35.9-99.6	90.2; 78.6-96.7	50.0; 18.7-81.3	97.9; 88.7-99.9
Nikolova, 2015 [62, 63]	CL<25mm	203	57.1; 39.4-73.7	72.6; 65.2-79.2	30.3; 19.6-42.9	89.1; 82.6-93.7
	fFN@50[d]	66	50.0; 21.1-78.9	72.2; 58.4-83.5	28.6; 11.3-52.2	86.7; 73.2-94.9
	PartoSure	203	80.0; 63.1-91.6	94.6; 90.1-97.5	75.7; 58.8-88.2	95.8; 91.5-98.3
Riboni, 2011 [56]	fFN@50[e]	210	50.0; 15.7-84.3	80.2; 74.0-85.5	9.1; 2.5-21.7	97.6; 93.9-99.3
	Actim Partus	210	50.0; 15.7-84.3	83.7; 77.8-88.5	10.8; 3.0-25.4	97.7; 94.2-99.4
Ting, 2007 [58]	fFN@50[f]	94	56.3; 29.9-80.2	75.6; 64.6-84.7	32.1; 15.9-52.4	89.4; 79.4-95.6
	Actim Partus	94	70.6; 44.0-89.7	77.9; 67.0-86.6	41.4; 23.5-61.1	92.3; 83.0-97.5
Tripathi, 2016 [59]	fFN@50[d]	468	23.8; 17.3-31.4	99.1; 97.3-99.8	92.3; 79.1-98.4	73.2; 68.7-77.3
	Actim Partus	467	94.7; 89.9-97.7	92.4; 88.9-95.1	85.7; 79.5-90.6	97.3; 94.8-98.8

Key: CL, cervical length; fFN, fetal fibronectin; n, number; NPV, negative predictive value; PPV, positive predictive value

Notes: a, other cut-off's available; b, raw data back calculated from sensitivity, specificity, PPV and NPV; c, quantitative Rapid fFN 10Q Cassette; d, QuikCheck fFN; e, fFN measured by ELISA; f, fFN testing method unclear

A3.1 Test accuracy data for transvaginal cervical length

A3.1.1 Quantity and quality of the data available for cervical length

As can be seen in Table 60, nine of the 20 included studies (APOSTEL-1, EUIFS, Azlin 2010, Bolotskikh 2017, Danti 2011, Eroglu 2007, Goyal 2016, Hadzi-Lega 2017, and Nikolova 2015) report diagnostic test accuracy data for cervical length (in addition to data for at least one index test).[45-47, 50, 52-54, 61-64]

Of these nine studies, four (APOSTEL-1, Eroglu 2007, EUIFS and Bolotskikh 2017) used the cervical length threshold recommended in the current NICE guidance (<15mm).[26, 45, 46, 53, 61, 64] Three of these studies also reported test accuracy data at other thresholds (not presented in this report).[45, 46, 61, 64] One study (Danti 2011) reported test accuracy of cervical length using the threshold of <20mm.[52] The remaining four studies (Azlin 2010, Goyal 2016, Hadzi-Lega 2017 and Nikolova 2014) all used a cervical length threshold of <25mm.[47, 50, 54, 62, 63]

One study (Danti 2011) reports cervical length test accuracy data for two populations: all recruited women had their cervical length measured. For those women with a cervical length >30mm, the Actim Partus test was not performed (n=42) and for those with a cervical length <30mm, Actim Partus was performed (n=60). Cervical length test accuracy data was available for both the women with a cervical length <30 (n=60) (i.e. for those women who also had an Actim Partus test) and for the whole sample (n=102).

It should be noted here that cervical length measurement is a more subjective test (i.e. more open to human interpretation) than any of the other tests (PartoSure, Actim Partus, quantitative fFN or qualitative fFN) and is, therefore, more dependent on clinicians' experience/expertise and more open to potential (intentional or unintentional) bias. Typically, it was reported that cervical length was measured by a trained investigator, and that three measurements were taken and averaged. However, it is generally unclear whether the clinicians measuring cervical length were blinded to the results of any biomedical test used. Indeed, with the exception of Eroglu (2007), all studies that evaluated cervical length did not clearly describe whether clinicians were blinded to other test results.[45-47, 50, 52, 54, 61-64] In Eroglu (2007) it was explicitly stated that the assessor was blinded to other test results.[53]

A3.1.2 Test accuracy of transvaginal cervical length measurement

Table 60 provides diagnostic test accuracy data for the three studies (APOSTEL-1, EUIFS and Bolotskikh 2017) assessing cervical length at a threshold of <15mm (against the 7 day delivery reference standard).[45, 46, 61, 64] At this threshold, sensitivity ranged widely, from 33.3% (95% CI 9.9, 65.1) in Bolotskikh (2017) to 72.5% (95% CI 60.4, 82.5) in APOSTEL-1.[45, 46, 61] Specificity was more similar across the four studies, ranging from 81.8% (95% CI 77.7, 85.4) in EUIFS to 98.9% (95% CI 93.8, 100.0) in Bolotskikh (2017).[61, 64]

Against the 7 day delivery reference standard, sensitivity of cervical length at a threshold of <20mm was 75.0% (95% CI 19.4, 99.4) in both of the Danti (2011) samples (n=60 and n=102) whilst specificity was 83.7% (95% CI 74.8, 90.4) for the sample of n=102 and 71.4% (95% CI 57.8, 82.7) in the sample of n=60 (i.e. women with a cervical length <30mm).[52] Eroglu (2007)[53] also evaluated cervical length at a threshold of <20mm against the 7 day delivery reference standard, sensitivity was lower 66.7 (95% CI 22.3, 95.7), however specificity was higher 95.6 (95% CI 84.9, 99.5) than Danti (2011). Across the four studies providing data at a threshold of <25mm, and again against a 7 day delivery reference standard, sensitivity ranged from 57.1% (95% CI 39.4, 73.7) in Nikolova (2015) to 100.0% (95% CI 54.1, 100.0) in Hadzi-Lega (2017).[47, 62, 63] In these four studies, specificity ranged from 31.6% (95% CI 12.6, 56.6) in Goyal 2016 to 72.6% (95% CI 65.2, 79.2) in Nikolova 2015.[54, 62, 63]

Again, it should be noted that these data do not cover all available evidence regarding test accuracy of cervical length at thresholds of <15mm, <20mm or <25mm and are based only on data reported by studies included in our systematic review of Actim Partus, PartoSure and quantitative fFN at thresholds other than 50ng/ml. In addition, the large variation across these studies in sensitivity and specificity may be, at least in part, due to the different clinical personnel conducting the cervical length measurements.

A3.1.3 Comparison of cervical length and index tests

Six studies (Hadzi-Lega 2017, APOSTEL-1, Azlin 2010, Eroglu 2007, Goyal 2016, Danti 2011) assessed both cervical length measurement and Actim Partus in the same population.[45-47, 50, 52-54] Three studies (Hazdi-Lega 2017, Bolotskikh 2017, Nikolova 2015) assessed both cervical length measurement and PartoSure,[47, 61-63] and two (APOSTEL-1, EUIFS) assessed both cervical length and quantitative fFN.[45, 46, 64] Note that both Hazdi-Lega (2017) and APOSTEL-1 assess two index tests.[45-47]

Against the 7 day reference standard, sensitivity was higher for cervical length measurement than Actim Partus in three studies (Hadzi-Lega 2017, Goyal 2016 and Danti 2011), see Table 60.[47, 52, 54] The cervical length threshold for a positive test result was <25mm in Hadzi-Lega (2017) and Goyal (2016), and <20mm in Danti (2011).[47, 52, 54] In one study (Azlin, 2010), sensitivity (against the 7 day reference standard) did not differ between Actim Partus and cervical length measurement with a threshold of <25mm.[50] In the remaining two studies (APOSTEL-1 and Eroglu 2007), sensitivity was higher for Actim Partus compared to cervical length measurement with a threshold of <15mm.[45, 46, 53] Specificity (against the 7 day reference standard) was higher for Actim Partus compared to cervical length in all studies except for Danti (2011) and Eroglu (2007), where the specificity was higher for cervical length (Table 60).[52, 53]

When comparing test accuracy of cervical length measurement with that of PartoSure (against the 7 day reference standard), sensitivity was higher for PartoSure than for cervical length at a threshold of <15mm in Bolotskikh (2017) (100%, 95% CI 73.5, 100.0 versus 33.3%, 95% CI 9.9, 65.1) and lower for PartoSure than for cervical length at a threshold of <25mm in Hadzi-Lega (2017) (83.3% 95% CI 35.9, 99.6 versus 100%, 95% CI 54.1, 100), although confidence intervals overlap in Hadzi-Lega (2017).[47, 61] Conversely, specificity was lower for PartoSure than for cervical length at a threshold of <15mm in Bolotskikh (2017) (95.4%, 95% CI 88.6, 98.7 versus 98.9%, 95% CI 93.8, 100.0) and higher for PartoSure than for cervical length at a threshold of <25mm in Hadzi-Lega (2017) (90.2%, 95% CI 78.6, 96.7 versus 70.6%, 95% CI 56.2, 82.5), albeit with overlapping confidence intervals.[47, 61] In the third study (Nikolova, 2015), both sensitivity and specificity were higher for PartoSure than

for cervical length at a threshold of <25mm (sensitivity 80.0%, 95% CI 63.1, 91.6 versus 57.1%, 95% CI 39.4, 73.7; specificity 94.6%, 95% CI 90.1, 97.5 versus 72.6%; 95% CI 65.2, 79.2), although again, confidence intervals (for sensitivity) overlap.

In comparison to quantitative fFN, in APOSTEL-1, cervical length at a threshold of <15mm was most closely matched to quantitative fFN with the threshold of 200ng/ml (sensitivity against the 7 day reference standard: 72.5%, 95% CI 60.4, 82.5 versus 71.0%, 95% CI 58.8, 81.3 respectively; specificity against the 7 day reference standard: 83.3%, 95% CI 78.4, 87.4 versus 83.6%, 95% CI 78.8, 87.8 respectively).[45, 46] However, in EUIFS, the sensitivity and specificity (against the 7 day reference standard) of cervical length measurement at a threshold of <15mm fell between sensitivities and specificities produced for quantitative fFN at the 200ng/ml and 500ng/ml thresholds (see Table 60).[64] This was particularly because sensitivity of cervical length at a threshold of <15mm was lower in EUIFS than in APOSTEL-1 (51.3%, 95% CI 34.8, 67.6 versus 72.5%, 95% CI 60.4, 82.5) although the 95% confidence intervals do overlap.[45, 46, 64]

A3.1.4 Data from systematic reviews of cervical length measurement

In the Boots et al. (2014) review, cervical length measurement at a cut-off of 15mm was assessed in 24 studies (against a reference standard of delivery within 7 days), with pooled sensitivity reported as 74% (95% CI 58, 85) and pooled specificity as 89% (95% CI 85, 92).[68] Recent NICE guidance shows how the variability across studies is great: for cervical length measurement at a cut-off of <15mm, across eight studies of 'very low' quality, sensitivity (against a reference standard of delivery within 7 days) ranged from 26.3% (95% CI 11.2, 39.7) to 97.7% (95% CI 86.9, 99.9) and specificity from 83.0% (95% CI 70.0, 93.0) to 96.5% (95% CI 95.4, 97.7).[26] These systematic review data are similar to the data for cervical length measurement from the current overview (see section A3.1.2 and Table 60) where, at a threshold of <15mm, sensitivity (against the 7 day reference standard) showed great variability across studies, ranging from 33.3% (95% CI 9.9, 65.1) to 72.5% (95% CI 60.4, 82.5), and specificity was more similar across studies, ranging from 81.8% (95% CI 77.7, 85.4) to 98.9% (95% CI 93.8, 100.0).

In the recent NICE guidance, at a cut-off of <25mm (across five studies of 'low' and 'very low' quality), sensitivity (against a reference standard of delivery within 7 days) ranged from 60.0% (95% CI 48.3, 64.7) to 83.3% (95% CI 43.7, 97.0) and specificity from 71.7% (95% CI 66.4, 73.8) to 96.9% (95% CI 91.6, 99.5).[26] Again, these sensitivity data are similar to those in the current overview where, at a threshold of <25mm, sensitivity ranged from 57.1% (95% CI 39.4, 73.7) to 100.0% (95% CI 54.1, 100.0). However, at this threshold, a wider range of specificity was found in the current overview (ranging from 31.6%, 95% CI 12.6, 56.6 to

72.6%, 95% CI 65.2, 79.2) compared with the recent NICE guidance.[26] The recent NICE guidance also included test accuracy data for cervical length at a threshold of <30mm (across three studies of 'very low' quality), with sensitivity (against a reference standard of delivery within 7 days) ranging from 89.3% (95% CI 71.8, 97.2) to 94.0% (95% CI 79.0, 99.0) and specificity from 42.0% (95% CI 37.0, 47.0) to 55.6% (95% CI 53.0, 56.8).[26] The current overview does not provide test accuracy data for cervical length at the <30mm threshold.

A3.1.5 Summary

Overall summary tables for the diagnostic test accuracy review, including CL data are presented in Table 61 and Table 62.

Table 58 Summary of evidence and relative accuracy against the 7 day reference standard

		Actim Partus	PartoSure	Quantitative fFN		
				@10ng/ml	@200ng/ml	@500ng/ml
Index tests						
Actim Partus						
PartoSure		No difference (Hadzi-Lega 2017)				
Quantitative fFN	**@10ng/ml**	Sensitivity of fFN superior, specificity of Actim Partus superior (APOSTEL-1)	Indirect evidence only			
	@200ng/ml	No difference (APOSTEL-1)	Indirect evidence only			
	@500ng/ml	Sensitivity of Actim Partus superior, specificity of fFN superior (APOSTEL-1)	Indirect evidence only			
fFN at 50ng/ml						
Quantitative fFN at 50ng/ml		Specificity of Actim Partus superior, no difference in sensitivity (APOSTEL-1)	Indirect evidence only	Sensitivity of fFN @10ng/ml superior, specificity of fFN @50ng/ml superior (APOSTEL-1 and EUIFS)	Sensitivity of fFN @50ng/ml superior, specificity of fFN @200ng/ml superior (APOSTEl-1 and EUIFS)	Sensitivity of fFN @50ng/ml superior, specificity of fFN @500ng/ml superior (APOSTEl-1 and EUIFS)
QuikCheck		Sensitivity of Actim Partus superior & specificity of fFN superior (Tripathi 2016). However Eroglu 2007 showed no difference between tests.	Specificity of PartoSure superior, no difference in sensitivity (Nikolova 2015-note missing participants)	Indirect evidence only	Indirect evidence only	Indirect evidence only
ELISA		No difference (Riboni)	No evidence	Indirect evidence only	Indirect evidence only	Indirect evidence only
Cervical length[a]						
CL<15mm		No difference (APOSTEL-1)	No evidence	Sensitivity of fFN superior, specificity of CL superior (APOSTEL-1 and EUIFS)	No difference (APOSTEL-1 and EUIFS)	Sensitivity of CL superior or no difference, specificity of fFN superior. (APOSTEL-1 and EUIFS)
CL<20 mm		No difference (Danti 2011 and Eroglu 2007)[b]	No evidence	Indirect evidence only	Indirect evidence only	Indirect evidence only
CL<25mm		No difference (Azlin 2010, Goyal 2016 and Hadzi-Lega 2017)	Specificity of PartoSure superior or no difference, Sensitivity no difference (Nikolova 2015 and Hadzi-Lega 2017)	Indirect evidence only	Indirect evidence only	Indirect evidence only

Key: **a**, Studies reporting accuracy of cervical length across multiple thresholds, data regarding the cut-off closest to 15mm threshold (NICE guidelines) is reported here. **b**, Danti 2011 a subset of recruited participants received Actim Partus test, no difference between CL and Actim Partus was observed when compared in the population that received both tests.

Table 59 Summary Table

Test Accuracy for the Prediction of Preterm Delivery within 7 days

Studies assessing more than one index test

Index Test	Source	Sensitivity % (95% CI)	Specificity % (95% CI)
fFN at 10ng/ml	Bruijn APOSTEL-1 (n=350)	95.7 (87.8, 99.1)	42.3 (36.5, 48.4)
fFN at 200ng/ml	Bruijn APOSTEL-1 (n=350)	71.0 (58.8, 81.3)	83.6 (78.8, 87.8)
fFN at 500ng/ml	Bruijn APOSTEL-1 (n=350)	42.0 (30.2, 54.5)	95.7 (92.7, 97.8)
Actim Partus	Bruijn APOSTEL-1 (n=350)	78.3 (66.7, 87.3)	89.3 (85.1, 92.7)
PartoSure	Hadzi-Lega 2017 (n=57)	83.3 (35.9, 99.6)	90.2 (78.6, 96.7)
Actim Partus	Hadzi-Lega 2017 (n=57)	83.3 (35.9, 99.6)	76.5 (62.5, 87.2)

Studies assessing a single index test

Index Test	Source	Sensitivity % (95% CI)	Specificity % (95% CI)
Actim Partus	Pooled (16 studies)	77 (68, 83)	81 (76, 85)
	Range (16 studies)	33.3 (4.3, 77.7) - 94.7 (89.9, 97.7)	50.0 (24.7, 75.3) - 93.5 (82.1, 98.6)
PartoSure	Pooled (4 studies)	83 (61, 94)	95 (89, 98)
	Range (4 studies)	0 (0.0, 97.5) - 100.0 (73.5, 100.0)	90.2(78.6, 96.7) - 97.5(96.8, 99.9)
fFN at 10ng/ml	Range (2 studies)	93.8 (82.8, 98.7) - 95.7 (87.8, 99.1)	32.2 (27.7, 37.0) - 42.3 (36.5, 48.4)
fFN at 200ng/ml	Range (2 studies)	70.8 (55.9, 83.0) - 71.0 (58.8, 81.3)	78.6 (74.3, 82.5) - 83.6 (78.8, 87.8)
fFN at 500ng/ml	Range (2 studies)	29.2 (17.0, 44.1) - 42.0 (30.2, 54.5)	94.3 (91.6, 96.4) - 95.7 (92.7, 97.8)

Supplementary data from included studies

Test	Source	Sensitivity % (95% CI)	Specificity % (95% CI)
fFN at 50ng/ml	Range (8 studies)	23.8 (17.3, 31.4) - 91.3 (82.0, 96.7)	62.2 (57.3,66.9) - 99.1 (97.3,99.8)
CL <15mm	Range (3 studies)	33.3 (9.9, 65.1) - 72.5 (60.4, 82.5)	81.8 (77.7, 85.4) - 98.9 (93.8, 100.0)
CL<20mm	Danti 2011 (n=60)	75.0 (19.4, 99.4)	71.4 (57.8, 82.7)
CL<25mm	Range (4 studies)	57.1 (39.4, 73.7) - 100.0 (54.1, 100.0)	31.6 (12.6,56.6) - 72.6 (65.2,79.2)

Data extracted from Systematic Reviews

Test	Source	Sensitivity % (95% CI)	Specificity % (95% CI)
fFN at 50ng/ml	Sanchez-Ramos 2009 Pooled (32 studies)	76.1 (69.1,81.9)	81.9 (78.9, 84.5)
fFN at 50ng/ml	Boots 2014 Pooled (38 studies)	75 (69, 80)	79 (76, 83)
fFN at 50ng/ml	NICE guidance Range (20 studies)	56[a] -100[a]	61.9 (59.6, 62.5) - 92[a]
CL<15mm	Boots 2014 Pooled (24 studies)	74 (58, 85)	89 (85,92)
CL<15mm	NICE guidance Range (8 studies)	26.3 (11.2, 39.7) - 97.7(86.9, 99.9)	83.0 (70.0, 93.0) - 96.5 (95.4, 97.7)
CL <25mm	NICE guidance Range (5 studies)	60.0(48.3, 64.7) - 83.3 (43.7, 97.0)	71.7 (66.4, 73.8) - 96.9 (91.6, 99.5)
CL <30mm	NICE guidance Range (3 studies)	89.3 (71.8, 97.2) - 94.0 (79.0, 99.0)	42.0 (37.0, 47.0) - 55.6 (53.0, 56.8)

Test Accuracy for the Prediction of Preterm Delivery within 48 hours

Studies assessing a single index test

Index Test	Source	Sensitivity % (95% CI)	Specificity % (95% CI)
Actim Partus	Pooled (6 studies)	87 (74, 96)	73 (62, 82)
	Range (6 studies)	65.7 (47.8, 80.9) – 100 (47.8, 100.0)	56.0 (34.9,75.6) - 82.4 (56.6, 96.2)
PartoSure	Werlen 2015 (n=41)	0.0 (0.0, 97.5)	97.5 (86.8, 99.9)

Key: [a] 95% CI not reported

Appendix 4. Supplementary discussion and tables for the systematic review and selection of utilities

A4.1 Quality of Life Outcomes for Preterm Children

Studies concerning the quality of life outcomes of preterm children are summarised in Table 60. Of the 24 papers shortlisted, 7 were deemed as lower priority, since they either use non-standard measures of quality of life, or do not report their quality of life figures in a usable format. One study does not report SF-36 mean scores, apart from in the form of a graph.[146] A second study measures but does not report any SF-36 scores.[112] Four studies use quality of life measures that do not have mapping functions that allow for conversion to utilities.[111, 147-149] Finally, another study measures utilities for 140 15-16 year-olds that had a gestational age below 29 weeks using the Health Utilities Index v3 (HUI3).[150] However, this was only an abstract, and it did not report any values.

Of the remaining 17 papers (authors marked as bold in Table 60), 12 provide direct measures of utility. Seven of these studies use a version of the HUI.[151-157]. One of these studies is a systematic review of quality of life, and also reports utilities drawn from other sources.[154]

Three studies from Finland use the 17D measure.[158-160] Five studies reported means and standard deviations for the SF-36 measure of health-related quality of life.[120-122, 125, 126] The remaining two studies are modelling papers that make use of utilities drawn from other sources.[118, 161]

Many studies in Table 60 that were model-based cited sources for the utilities they used. These source papers were collated and summarised in Table 61. The majority of these studies used either: the HUI2 domains, which had been converted into a utility using a multi-attribute health status utility function[162]; or a direct utility measure based on standard gamble. The study by Carroll and Downs and some of the studies reported in Tengs and Wallace also used TTO methods of utility elicitation.[117, 163]

Table 60 Summary of papers from our systematic search with information on quality of life for preterm children

Paper	Population	Sample size	Country	QoL measure	Parameters provided	Comments
Bastek et al. 2012[118]	Preterm children with 34 ≤ gestational age (GA) < 36 weeks	N/A (literature review)	USA (though utilities obtained from other studies)	Standard gamble and time trade-off methods used in source paper.	Utilities for acute respiratory disease; chronic respiratory disease; neurodevelopmental delay in childhood; death in childhood.	These utilities originate from two sources.[117, 124]. Utilities for moderate persistent asthma and moderate cerebral palsy were used as proxies for RDS and adverse neurodevelopment, respectively.
Batsvik et al. 2015[121]	GA ≤ 28 weeks or birthweight (BW) ≤ 1000g; assessed at mean age of 24; with/without severe disability	43 preterm + 43 control	Norway	SF-36	SF-36 dimension means	
Baumgardt et al. 2012[146]	Preterm with BW < 1250g; surveyed at median age of 23	52 preterm + 75 control	Switzerland	SF-36	SF-36 dimension means plotted but not explicitly provided.	Scores are also separated by sex
Beaudoin et al. 2013[112]	Preterm with bronchopulmonary dysplasia (BPD); with RDS; with no respiratory complications	426 with BPD + 852 RDS / preterm / term	Canada	SF-36 v2	SF-36 results not reported	
Berbis et al. 2012[147]	Gestational age between 24 and 32 weeks; assessed at age 6-10	82	France	VSP-A	VSP-A subscale means reported in paper, but would need to be combined to form a utility	Preterm children are compared to French population norms
Bianco et al. 2011[148]	GA ≤ 29 weeks and/or BW ≤ 1500g; treated/not treated with surfactant; assessed at 18 months	89 preterm w/ surfactant + 61 preterm, no surfactant + 145 term	Italy	TAPQOL	TAPQOL of children treated / not treated with surfactant.	Abstract only
Cooke et al. 2004[120]	Preterm very low birthweight (VLBW) infants, assessed at age 19-22 (mean 20)	79 preterm + 71 term	UK	SF-36	SF-36 dimension means reported for both males and females	The paper also reports additional information on social/behavioural outcomes, depression, and physical size
Dalziel et al. 2007[125]	Preterm and term children, assessed at age 31	126 preterm + 66 term	New Zealand	SF-36	SF-36 dimension means	

Paper	Population	Sample size	Country	QoL measure	Parameters provided	Comments
Einerson et al. 2002[161]	N/A – paper presents a cost-effectiveness model for cervical length screening	N/A – see source paper information in Table 61.	USA (though utilities obtained from other studies)	Standard gamble and time trade-off methods used in source papers.	Utilities for neonatal death; severe neonatal morbidity; healthy neonate.	Utilities for death and morbidity obtained from three sources.[124, 131, 163] It is not clear exactly which figures have been used, or how they may have been combined.
Feingold et al. 2002[111]	BW < 1501g and GA < 33 weeks, assessed at age 18-19	43 IVH 0-2, no cysts + 10 IVH 3-4 and / or cysts	USA	Health-Related Quality of Life (HRQL) from CDC	Means of 4 dimensions of HRQL reported, separated into 2 IVH severity level groups (0-2, no periventricular leukomalacia (PVL); and 3-4, with/without PVL).	Unclear how to derive utilities from the HRQL means
Gray et al. 2007[150]	GA < 29 weeks, assessed at age 15-16	140 preterm + 108 control	UK	HUI3	No figures provided	Abstract only - relative differences are reported but not absolute utilities.
Husby et al. 2016[126]	Preterm with BW ≤ 1500g, assessed at age 23	35 preterm + 37 control	Norway	SF-36	SF-36 dimension means provided for those VLBW children without cerebral palsy or low IQs, as well as those with one or more of the above.	Additional results find lower risk of alcohol abuse, 5 times higher likelihood of depression, and poorer motor skills.
Ketharanathan et al. 2011[149]	32 ≤ GA < 36 weeks, assessed at pre-school age (2-5 years)	218 responders	Netherlands	TAPQOL	TAPQOL dimension means are provided	Unclear how to derive utilities from the TAPQOL means. Study also reports prevalence of various behavioural problems.
Korvenranta et al. 2010[158]	GA < 32 weeks or BW < 1501g, assessed at age 4	1752	Finland	Utilities derived from 17D[160]	Implied utilities provided for survivors (QALYs/4). These include survival utilities for different gestational ages; seizures; cerebral palsy; visual problems; hearing problems; obstructive airway diseases.	QALYs were calculated for 4 years in the paper by defining a HRQOL score for each day of life, then multiplying by number of days alive
Lehtonen et al. 2011[159]	GA < 32 weeks or BW < 1501g, assessed at age 5	568 preterm + 173 control	Finland	Utilities derived from 17D[160]	Implied utilities (QALYs/5) for preterm/VLBW compared with controls.	Being born in a level III hospital increased median QALY by 0.03/5 = 0.006, relative to a level II hospital.

Paper	Population	Sample size	Country	QoL measure	Parameters provided	Comments
Lund et al. 2012[122]	Preterm with BW ≤ 1500g, and another group small for gestational age (SGA). Assessment at age 20	43 VLBW + 55 SGA + 73 control	Norway	SF-36	SF-36 dimension means provided for each group	Many other cognitive and behavioural measures also reported
Petrou et al. 2009[151]	20 ≤ GA < 25 weeks, assessed at age 11 (EPICure study).	190 preterm + 141 control	UK and Ireland	HUI3	HUI3 multi-attribute utilities provided for gestational ages up to 25 weeks, as well as controls.	Utility score for GA ≤ 23 weeks based on a sample of only 19. HUI3 scores were converted into multi-attribute utilities using methods from two studies.[164, 165].
Rautava et al. 2009[160]	BW ≤ 1500g or GA < 32 weeks, assessed at age 5	588 preterm + 176 control	Finland	17D	Figures identical to those used in Lehtonen et al.[159]. Additional utility provided for live-born VLBW.	Being born in a level III hospital (relative to a level II hospital) increased the mean QALY by 0.5/5 = 0.1.
Roberts et al. 2013[152]	GA < 28 weeks or BW < 1000g, assessed at age 18	194 preterm + 148 control	Australia	HUI3 and SF-36	HUI3 scores for preterm and controls. SF-36 dimension scores also provided, but only medians.	It is not clear whether the HUI3 score reported here is computed in the same way as the multi-attribute score in Petrou et al.[151].
van Lunenburg et al. 2013[153]	GA < 32 weeks or BW < 1500g, assessed at ages 19 and 28 (POPS cohort)	314	Netherlands	HUI3	HUI3 multi-attribute utility given at age 19, and at 28.	Multiple imputation values are based on an algorithm (multivariate imputation by chained equations) that incorporates information from respondents to interpolate missing data values
Verrips et al. 2008[154]	BW ≤ 1000g included from 3 separate cohorts, assessed at age 12-16	150 (Canada), 65 (Germany), 126 (Netherlands)	Germany, Canada, Netherlands	HUI3	HUI3 multi-attribute utility for ELBW in Canada, Germany, and Netherlands.	Utility function based on Furlong et al.[165]
Wolke 2016[155]	Extremely low birth-weight (ELBW, Canada); VP or VLBW (Germany, Netherlands), assessed at adolescence (12-16), early	169 (Canada), 91 (Germany), 140 (Netherlands)	Germany, Canada, Netherlands	HUI3	HUI3 utilities reported for 3 life stages for Canada, Germany, and Netherlands. Canada also includes utilities with/without neurosensory impairment.	Summary of utilities from multiple studies. Dutch study had a different cohort at adolescence, but the same cohort measured at both early adulthood and adulthood.

Paper	Population	Sample size	Country	QoL measure	Parameters provided	Comments
	adulthood (19-26), adulthood (>26)					
Wolke et al. 2013[156]	BW < 1500g or GA < 32 weeks, assessed at age 13	260 preterm + 282 control	Germany	HUI3	HUI3 multi-attribute utilities for VP/VLBW and controls, reported by both parents and children.	Paper also reports other social and cognitive characteristics. In particular, mean IQ in VP/VLBW = 92, versus 101 in full-term controls. Another group of VP/VLBW who could not report their own utility had a parent-reported value of 0.18, but only based on a small sample (n=12).
Zwicker and Harris 2008[157]	VLBW or preterm	N/A, see Table 61 for source paper information	Multiple, but utility sources all from Canada	HUI2 and standard gamble	Utilities for preterm and control children at school-age; adolescence; and young adulthood.	Study is a systematic review of quality of life scores from other studies. Utility scores reported are from multiple studies. [166-170]

Note: Rows shaded grey and fully in italics represent studies that did not provide any appropriate quality of life utility measures.

Table 61 Secondary sources for utilities (referenced by papers found in the systematic search)

Study	Sample	Country	Utility measure	Utilities reported	Comments
Carroll and Downs 2009[117]	4016 (over 18, had at least 1 child)	USA	Standard gamble and time trade-off	Utilities for 30 medical states provided (including perfect health).	Implicitly uses Von-Neumann-Morgenstern expected utility assumptions. Both elicitation methods assume risk neutrality.
Pham and Crowther 2003[131]	180 (90 postnatal, 59 midwives, 31 medical staff)	Australia	Standard gamble	Perfect health; Jaundice requiring phototherapy; Admission to neonatal nursery; Shoulder dystocia; Nerve palsy; Transient neurological symptoms; Permanent neurological sequelae; perinatal death	Health outcomes do not seem relevant to current model. Mothers of premature babies were excluded from the study.
Saigal et al. 1994[166]	156 ELBW + 145 controls	Canada	HUI2 transformed using a multi-attribute health status (MAHS) utility function	Utilities are provided for different attribute score combinations on a 6 dimension quality of life scale, for children at age 8.	Originally excluded in full text screening as it focused on low birthweight only.

Study	Sample	Country	Utility measure	Utilities reported	Comments
					Utilities provided are for a combination of subjective health states, and do not correspond to any particular condition.
Saigal et al. 1994b[168]	156 ELBW + 145 controls	Canada	HUI2 transformed using a multi-attribute health status (MAHS) utility function	Study identifies unique health states required to classify the ELBW and control children, but does not report utilities explicitly.	This study appears to be supplemental to the previous study in this table, rather than providing new utility data.
Saigal et al. 1996[167]	141 ELBW + 145 controls	Canada	HUI2 (actual and for hypothetical states), Standard gamble	Mean utilities for age 12-16 reported, for own health states as well as 4 hypothetical scenarios	Sensitivity analysis results also provided. Provides average utility for ELBW, but not for specific medical conditions in isolation.
Saigal et al. 2000[169]	149 (parents of ELBW) + 126 (parents of controls)	Canada	HUI2 (actual states), Visual-analog scale and standard gamble (hypothetical states)	Parental perspectives of child's (12-16 years) utility, and perspectives for 4 hypothetical scenarios	Originally excluded in full text screening as it focused on low birthweight only. Data for child's impression exists, so parent's assessment may be unnecessary.
Saigal et al. 2006[170]	143 ELBW + 130 controls	Canada	Standard gamble (quality of life, and hypothetical states)	Utilities for young adults (~23 years) with/without neurosensory impairments	Originally excluded in full text screening as it focused on low birthweight only. Results of sensitivity analyses also provided.
Tengs and Wallace 2000[163]	154 (studies reviewed)	Multiple	51% of studies used direct elicitation, 32% estimated, 17%, health status instruments	1000 health states reported. Relevant utilities are outcomes from various degrees of low birthweight, at different levels of severity.	Paper is a review of studies containing original quality of life estimates for 1000 health states. Would need to refer to original studies to critique individual utilities.
Vandenbussche et al. 1999[124]	42 (12 obstetricians, 15 pregnant women, 15 mothers)	Netherlands	Standard gamble	4 health states: healthy child; transient neurological symptoms; permanent neurological symptoms; neonatal death. Each outcome has 3 utilities depending on type of birth.	Sample size is split into pregnant women, mothers, and obstetricians. This does not consider longer term outcomes reported by preterm survivors.

A4.2 Quality of Life Outcomes for Mothers

Studies concerning the quality of life outcomes of mothers are summarised in Table 62. Of these four studies, one was not used since its measure of quality of life cannot be mapped into a utility value.[114] The remaining three studies report summary scores for the SF-36.[113, 115, 171]

Table 62 Summary of papers on quality of life for mothers

Paper	Population	Sample size	Country	QoL measure	Parameters provided	Comments
Alemdaroglu et al. 2009[113]	Mothers of LBW, premature children with/without ICH or IVH.	24 (12 with ICH/IVH children, 12 without)	Turkey	SF-36	QoL for mothers of children with and without ICH/IVH	Only abstract available, no further details of SF-36 dimensions. Sample size very small.
Couto et al. 2009 [115]	Pregnant women with a history of one or more of the following: recurrent abortion, fetal death, preterm birth, early neonatal death	120 prior adverse outcomes + 120 controls	Brazil	SF-36	SF-36 dimension means	May not be as relevant for mothers who are not likely to have more children
Coyle 2011[171]	Random sample of mothers of students in each of 4 different age groups (<5, 5-10, 11-13, 14-18)	234	USA	SF-36 v2	SF-36 dimension means	Also provides mean SF-36 measures for mothers after splitting them into 3 age groups
Hill and Aldag 2007[114]	Mothers of preterm, near-term and term children. Assessed 7 and 21 days post-delivery	184 (37 preterm, 59 near term, 88 term)	USA	MAPP-QOL	QoL for preterm, near term and term, evaluated 1 and 3 weeks after birth.	Mean scores for additional sub-dimensions also reported. Unclear how to convert this measure to a utility.

Note: Rows shaded grey represent studies that did not provide suitable quality of life information for the PenTAG model

A4.3 Summary tables containing raw and mapped utilities from all relevant studies identified in the systematic review of utilities

Table 63 Summary of utilities for preterm children

Study	Population	Measure	Variable	Utility
Bastek et al. 2012[118]	34 ≤ GA < 36	Standard gamble, TTO	Acute respiratory disease	0.87
			Chronic respiratory disease	0.88
			Neurodevelopmental delay	0.76
			Death	0.01
Batsvik et al. 2015[121]	GA ≤ 28 weeks or BW ≤ 1000g; 43 preterm + 43 control	SF-36 (assessed at mean age 24)	Severe disability	0.763
			Healthy	0.846
Cooke 2004[120]	Preterm (VLBW) infants;	SF-36 (assessed at age 19-22)	All preterm	0.879
			Male	0.907

Study	Population	Measure	Variable	Utility
	79 preterm + 71 control		Female	0.856
Dalziel et al. 2007[125]	Preterm and term children; 126 preterm + 66 control	SF-36 (assessed at age 31)	All preterm	0.887
Einerson et al. 2016[161]	Preterm survivors	Standard gamble, TTO	Neonatal death	0
			Severe neonatal morbidity	0.55
			Healthy neonate	1
Husby et al. 2016[126]	Preterm with BW ≤ 1500g; 35 preterm + 37 control	SF-36 (assessed at age 23)	All preterm	0.857
			VLBW with no cerebral palsy, and not with low IQ	0.891
Korvenranta et al. 2010[158]	GA < 32 or BW < 1501g; n=1752	17D (assessed at age 4)	None of the studied morbidities	0.9475
			23 weeks GA	0.9025
			24-25 weeks	0.9075
			26-27 weeks	0.9175
			28-29 weeks	0.9275
			30-31 weeks	0.94
			≥ 32 weeks	0.9425
			Seizures	0.9675
			Cerebral palsy	0.9225
			Visual disorder	0.875
			Other ophthalmologic problems	0.9375
			Hearing loss	0.8825
			Obstructive airway diseases	0.91
			2 or more of the above morbidities	0.87
Lehtonen et al. 2011[159] *and Rautava et al. 2009*[160]	GA < 32 or BW < 1501g; 568 preterm + 173 control	17D (assessed at age 5)	All preterm	0.92 (median) 0.72 (mean)
			Live-born	0.94 (median) 0.82 (mean)
Lund et al. 2012[122]	Preterm with BW ≤ 1500g, and small for gestational age (SGA); 43 VLBW + 55 SGA + 73 control	SF-36 (assessed at age 20)	All preterm	0.901
			SGA	0.888
Petrou et al. 2009[151]	20 ≤ GA < 25; 190 preterm + 141 control	HUI3 (assessed at age 11)	All preterm	0.789
			≤ 23 weeks	0.772
			24 weeks	0.717
			25 weeks	0.83
Roberts et al. 2013[152]	GA < 28 or BW < 1000g; 194 preterm + 148 control	HUI3 (assessed at age 18)	All preterm	0.93
van Lunenburg et al. 2013[153]	GA < 32 or BW < 1500g; n=314	HUI3	Age 19 (assessed)	0.89
			Age 19 (multiple imputation)	0.83
			Age 28 (assessed)	0.88
			Age 28 (multiple imputation)	0.85
Verrips et al. 2008[154]	BW ≤ 1000g; 150 (Canada), 65 (Germany), 126 (Netherlands)	HUI3 (assessed at 12-16)	Canada	0.76
			Germany	0.752
			Netherlands	0.868
Wolke 2016[155]	ELBW (Canada, n=169); VP or	HUI3, assessed at adolescence	Canada, all preterm	0.79 (12-16y) 0.79 (19-26y)

Study	Population	Measure	Variable	Utility
	VLBW (Germany n=91, Netherlands n=140)	(12-16), early adulthood (19-26), adulthood (>26)		0.73 (> 26y)
			ELBW without neurosensory impairment (Canada)	0.83 (12-16y) 0.83 (19-26y) 0.77 (> 26y)
			ELBW with neurosensory impairment (Canada)	0.68 (12-16y) 0.65 (19-26y) 0.60 (> 26y)
			Germany, all preterm	0.82 (12-16y) 0.82 (19-26y)
			Netherlands, assessed	0.87 (12-16y) 0.89 (19-26y) 0.88 (> 26y)
			Netherlands, imputed	0.83 (19-26y) 0.85 (> 26y)
Wolke et al. 2013[156]	GA < 32 or BW < 1500g; 260 preterm + 282 control	HUI3 (assessed at age 13)	Parent reported	0.88
			Child self-reported	0.84
Zwicker and Harris 2008[157]	VLBW or preterm	HUI2 and standard gamble	School-age	0.82
			Adolescents (study 1)	0.87
			Adolescents (study 2)	0.91
			Young adults	0.85

Table 64 Summary of utilities from secondary sources cited by studies found in the systematic search

Study	Population	Measure	Variable	Utilities	
Carroll and Downs 2009[117]	4016 (parent assessment of child's health)	Standard gamble and time trade-off	Mild persistent asthma	0.90 (SG)	0.91 (TTO)
			Mild intermittent asthma	0.91 (SG)	0.91 (TTO)
			Moderate persistent asthma	0.88 (SG)	0.91 (TTO)
			Severe persistent asthma	0.83 (SG)	0.85 (TTO)
			Mild cerebral palsy	0.87 (SG)	0.88 (TTO)
			10 day ICU hospitalisation	0.87 (SG)	0.91 (TTO)
			Mild seizure disorder	0.85 (SG)	0.86 (TTO)
			Moderate seizure disorder	0.84 (SG)	0.83 (TTO)
			Mild mental retardation	0.84 (SG)	0.83 (TTO)
			Moderate cerebral palsy	0.76 (SG)	0.76 (TTO)
			Severe seizure disorder	0.70 (SG)	0.71 (TTO)
			Severe cerebral palsy	0.60 (SG)	0.55 (TTO)
			Severe mental retardation	0.59 (SG)	0.51 (TTO)
Pham and Crowther 2003[131]	180 (90 postnatal, 59 midwives, 31 medical staff)	Standard gamble, median scores	Admission to neonatal nursery	0.99 (mothers)	0.95 (midwives) 0.99 (medical staff)
			Transient neurological symptoms	0.95 (mothers)	0.90 (midwives) 0.95 (medical staff)
			Permanent neurological sequelae	0.50 (mothers)	0.50 (midwives) 0.50 (medical staff)
Saigal et al. 1994[166]	156 ELBW + 145 controls	HUI2 transformed using a multi-attribute health status (MAHS) utility function	All ELBW	0.82 (mean)	0.88 (median)

Study	Population	Measure	Variable	Utilities	
Saigal et al. 1996[167]	141 ELBW + 145 controls	Standard gamble (chance board)	All ELBW	0.87 (mean)	1.00 (median)
Saigal et al. 2000[169]	149 (parents of ELBW) + 126 (parents of controls)	Standard gamble (chance board)	All ELBW (parental assessments)	0.91 (mean)	1.0 (median)
Saigal et al. 2006[170]	149 (parents of ELBW) + 126 (parents of controls)	Standard gamble	All ELBW	0.85 (mean, ELBW)	0.95 (median, ELBW)
			ELBW with neurosensory impairments	0.85	
			ELBW without neurosensory impairments	0.85	
Tengs and Wallace 2000[163]	140 experts	Standard gamble	Cerebrovascular disease, intracranial aneurysm, good but incomplete recovery, normal life with minor neurologic and psychological deficits	0.85	
			Cerebrovascular disease, intracranial aneurysm, moderate disability, independent daily living	0.63	
			Cerebrovascular disease, intracranial aneurysm, persistent vegetative state, unresponsive and speechless until death after acute brain damage	0.08	
			Cerebrovascular disease, intracranial aneurysm, severe disability, dependent on others for daily support due to mental and/or physical disability	0.26	
	156 patient proxies	HUI	ELBW (501-1000g), assessed at age 8	0.82	
	24 patients	Standard gamble	Lung disease, chronic, eg, chronic bronchitis, emphysema, cystic fibrosis, on waiting list for transplant	0.65	
			Lung disease, chronic, eg, chronic bronchitis, emphysema, cystic fibrosis, transplant	0.8	
Vandenbussche et al. 1999[124]	42 (12 obstetricians, 15 pregnant women, 15 mothers)	Standard gamble, median utilities	Healthy child, spontaneous birth	1	
			Transient neurological symptoms, spontaneous birth	0.99	
			Permanent neurological symptoms, spontaneous birth	0.5 (pregnant women) 0.35 (mothers)	0.05 (obstetricians)
			Neonatal death, spontaneous birth	0.01 (pregnant women, mothers)	0.23 (obstetricians)

Table 65 Mapped utilities for the quality of life of mothers

Study	Population	n	Measure	Variable	Utilities
Alemdaroglu et al. 2015[113]	Mothers of LBW, premature children with/without ICH or IVH.	12	SF-36 (physical and mental summary only)	Mothers of children with ICH/IVH	1.021†
		12		Mothers of children without ICH/IVH	1.016†
Couto et al. 2009[115]	Pregnant women with a history of one or more of the following: recurrent abortion, fetal death, preterm birth, early neonatal death	120	SF-36	Pregnant women with previous adverse pregnancy outcomes	0.644
		120		Control mothers	0.834
Coyle 2011[171]	Random sample of mothers of students in each of 4 different age groups (<5, 5-10, 11-13, 14-18)	234	SF-36 v2	All mothers	0.640
		69		Mothers 25-34 years	0.529*
		110		Mothers 35-44 years	0.525*
		40		Mothers 45-54 years	0.532*

Notes: * Utility mapped using linear model (rather than quadratic), since standard deviations were not reported for the input SF-36 mean scores. † These mapped utilities are greater than one, suggesting that the mapping function is extrapolating too far outside of the domain of the original sample used in Kim et al.[172]

A4.4 SF-36 mapping and extraction of utilities

Though none of the studies that were found directly measured utilities based on the EQ-5D, various mapping functions exist which allow SF-36 summary measures to be converted into EQ-5D utilities. The EQ-5D is NICE's preferred measure of quality of life.[123] The Oxford Health Economics Research centre maintains a database of such mapping studies.[173] The latest version was last updated in May 2016.

Two studies from this database were shortlisted, due to their use of a more general sample of the population, and a large sample size. The first uses UK data (n=25,783) and a generalised least squares (GLS) approach to estimate a mapping function to the EQ-5D.[110] They show that this provides a more accurate prediction of EQ-5D utility than using ordinary least squares (OLS) estimates. Whilst censored models were also estimated, these are problematic to use due to their non-linearity. Since only mean SF-36 data is provided in the papers included after screening, and given that the mean is a linear operator, it would not be possible to generate predictions without bias in censored mapping models. However, given that the studies provided standard deviations for their SF-36 summary scores, one can use mean-aggregated data to predict EQ-5D using the quadratic version of their GLS mapping

model.[3] This version of the model was preferred to the linear model where appropriate, as it provided an improved R^2 value (0.70 for the quadratic model, 0.67 for the linear model), and less than or equal mean squared errors everywhere outside the range 0-0.499.[110]

The second mapping study is not preferred to the first as it uses a smaller Korean sample (n=1660) to generate model estimates, which may be less representative of the UK population.[172] However, it includes a simple linear model (estimated using OLS) that generates EQ-5D utility from the physical health and mental health summary scores that are sometimes reported from SF-36 data. This model (R^2 = 0.6366, RMSE = 0.16) was used to predict EQ-5D from the single study that reported outcomes for mothers of children with and without IVH or ICH, since this study did not report mean scores for each of the eight SF-36 dimensions.[113]

These SF-36 to EQ-5D mapped utilities are tabulated in the Appendix, along with the other relevant utilities extracted directly from papers (where available) as follows. Table 63 contains all relevant utilities from studies on preterm children. Table 64 contains all relevant utilities from studies that were identified as secondary sources of utility data. Finally, Table 65 contains all relevant utilities from the studies on mothers.

A4.5 Study discussion

Studies on preterm children

The majority of the studies identified utilities (either directly or via the SF-36) for children born preterm or at a reduced birthweight, at various stages of life. Of these, only two studies used a UK/Ireland based population.[120, 151]

Two studies provide utilities for children born at different gestational ages.[151, 158] Petrou et al. only study children born between the gestational ages of 20 and 25 weeks, but assess utility using the HUI3 at 11 years of age on a UK and Ireland population. [151] Korvenranta et al. study Finnish children at 4 years of age, but provide utilities using the 17D measure for all gestational ages of preterm birth from 23 weeks onwards.[158]

The largest studies are from Finland, and all three make use of the 17D quality of life measure.[158-160] These studies report QALYs rather than utilities, but since they are computed linearly, implied utilities are derived by dividing the QALY value by the overall time horizon (4 years in Korvenranta et al.; 5 years in Lehtonen et al).[158, 159] No mapping studies from 17D to EQ-5D were found in the current Oxford Health Economics Research centre database.[173] A

[3] The study also includes a version of the model with full interaction terms. However, this was not used as it provided only an incremental improvement in fit, whilst introducing bias. This is due to the assumption required that covariances between SF-36 dimension means are 0, given that the studies that report SF-36 means did not, in general, report a full variance-covariance matrix.

Google Scholar search for the term ’17D EQ-5D’ was undertaken, but no mapping studies between the two were found.

Only one paper considers the quality of life for preterm children with IVH, separated into two severity groups: level 0-2 IVH with no PVL; and level 2-4 IVH with/without PVL.[111] The health-related quality of life measure they used, developed by the Centers for Disease Control and Prevention (CDC), does not have a suitable mapping to EQ-5D utility. However, those with more severe IVH do, in most cases, report a significantly lower quality of life at age 18-19.

Likewise, only one paper considers quality of life for preterm children with RDS. They measure, but do not provide, SF-36 scores in their paper.[112] However, they do explain that there was no significant difference in SF-36 means between different groups of children, when assessed as adults. This study is discussed further in A4.6.

Whilst other studies do not measure quality of life for children with IVH or RDS, some do make use of related utility measures as proxies for these conditions. The best example of this [118] identifies utilities (originally derived from two other studies [117, 124]) for acute/chronic respiratory disease, and neurodevelopmental delay. The utilities from Carroll and Downs are considered particularly reliable, as they are the result of using both the standard gamble and the time trade-off methods of elicitation for 4016 US parent (or guardian) assessments of a child’s hypothetical health state.[117] The utilities from this paper are used by Bastek et al. for their model of antenatal corticosteroids, a treatment which is relevant to the economic model devised for this report.[118] This study is discussed further in A4.6 and A4.7.

We received a forthcoming paper in confidence after contacting Dr Stavros Petrou, a health economist who has previously studied childhood outcomes.[174] This study includes a meta-analysis of utilities for preterm birth, as well as for other complications that may be related to RDS and IVH. This may provide more reliable quality of life estimates than selecting one study alone, but has the disadvantage of only providing utilities classified into more general health states than those found in individual studies. For example, they provide a weighted average utility score for chronic lower respiratory disease, and for combined disorders of the respiratory system, but not specific utilities for RDS. In addition, there are no average utilities for children born preterm measured by EQ-5D or SF-36/SF-6D. Finally, this study excluded papers with a mean or median assessment age higher than 18, which may be problematic when extrapolating over the entire lifespan.

Studies on mothers

The quality of evidence on the quality of life of mothers of preterm children is low. Only two studies consider mothers of preterm children specifically.[113, 114] The first is an abstract that reports only physical and mental health SF-36 mean summary scores. It is taken from a small Turkish sample of 24 mothers (12 who have LBW preterm children with ICH or IVH, and 12 who have LBW preterm children without ICH or IVH), which may not be representative of mothers in the UK. Furthermore, given that the mapping function (derived from subjects in Korea) applicable to physical and mental summary SF-36 scores used OLS,[172] it yields utilities greater than 1 (see Table 65) when applied to the data from Alemdaroglu et al.[113] Hence, we are not able to use this paper to generate appropriate utilities for the economic model.

The second study reported MAPP-QOL scores for mothers in the USA.[114] This study could not be mapped into utilities, and only provides quality of life for mothers of preterm, near-term, and term children 1 and 3 weeks postpartum. However, it does not contain information on mothers of preterm children with adverse birth conditions. Likewise, the utility mapped from the SF-36 means of a random sample of US mothers by Coyle could have been considered as a baseline for the quality of life of mothers whose children do not experience adverse health outcomes, but there is no corresponding quality of life information for mothers of children who have adverse health outcomes.[171] Therefore, neither of these studies provide usable utilities for the economic model.

The final study on mothers, by Couto et al., captures the quality of life in mothers in Brazil who have had at least one of four previous adverse pregnancy outcomes.[115] Whilst preterm birth is one of the four outcomes that is an inclusion criterion (along with early neonatal death, recurrent abortion, and fetal death), we are not provided with separate utilities for each outcome individually. The death outcomes are likely to skew the utility measure lower than if only mothers with a history of preterm birth were included in the population. It may be useful to treat this utility as a proxy for any adverse outcomes resulting from preterm labour, with the caveat that the utility would be an underestimate for mothers of preterm children that develop conditions but do not die; and an overestimate for mothers of preterm children that die.

A recent study (that was not included in the shortlist) suggests that whether a child is born very preterm or not may not have much of an effect on longer term parent quality of life.[175] We also consulted with Professor Dieter Wolke, an expert on the outcomes of preterm and low birthweight children, on this matter. He confirmed that, on the whole, data on the outcomes of parents with preterm children is very limited.

A4.6 Utilities for Respiratory Distress Syndrome (RDS)

There is only one study identified that measured quality of life outcomes for preterm children with RDS.[112] The study compares four groups of subjects:

1. Subjects born preterm (gestational age < 37 weeks) who developed bronchopulmonary dysplasia (BPD) without infant RDS
2. Subjects born preterm with RDS but no subsequent BPD
3. Subjects born preterm without respiratory complications
4. Subjects born at term without respiratory complications

The study uses a Canadian sample that were administered questionnaires containing the SF-36v2 via mail. 233 of the responses were from preterm individuals with RDS, measured at a mean age of 20.04 years. However, SF-36 scores were not reported in the paper. The corresponding author was contacted to request this data, but was unable to provide it. The study claims that health-related quality of life did not differ between the four groups studied.

In their modelling study for antenatal corticosteroids, Bastek et al. utilise proxy utilities for acute and chronic respiratory disease.[118] They argue that a utility value for a 10-day ICU admission was an acceptable proxy for acute respiratory disease, since infants with RDS are managed in NICUs. The utility corresponding to this outcome is 0.87.[117] From Table 64, we see that Carroll and Downs obtained this utility using the standard gamble method. Since NICE uses the UK time trade-off (TTO) value set to obtain utilities from the EQ-5D, TTO utilities obtained by Carroll and Downs are preferred to those obtained by standard gamble.[123] Carroll and Downs' TTO utility for 10-day ICU admission is 0.91.[117]

Bastek et al. report a utility for chronic respiratory disease of 0.88, which was taken from Carroll and Downs as the utility of moderate persistent asthma.[117, 118] From Table 64, we see that this utility was obtained using the standard gamble method. The TTO equivalent from Carroll and Downs is 0.91.[117] However, given that this utility is higher than the value used for preterm survivors, we opted to use the utility for severe persistent asthma from Carroll and Downs. This is 0.85, as elicited using the TTO method.

A4.7 Utilities for Intraventricular Haemmorhage (IVH)

As discussed in A4.5, only one study measured quality of life outcomes for children specifically with IVH.[111] This study uses four health-related quality of life questions from the CDC. One of these is measured on a 5-point scale, whilst the remaining three ask for a number of days over the past 30 days that a particular health state (e.g. poor mental health)

was not good. There is no clear mapping for these measures to provide a single measure of utility.

Therefore, the study by Bastek et al. again provides the best available estimate of a utility value.[118] They use the utility for moderate cerebral palsy from Carroll and Downs as a proxy for adverse neurodevelopment.[117] From Table 64, we see that this utility of 0.76 is identical for both the standard gamble and TTO methods of elicitation.

A4.8 Utilities for preterm survivors

NICE's preferred measure of health-related quality of life in adults is the EQ-5D.[123] However, given that we do not have any such data, the second-best option is to use SF-36 scores mapped onto EQ-5D. Five studies provide utilities for children that were obtained from mapping the SF-36 mean dimension scores onto the EQ-5D.[120-122, 125, 126] Whilst many other studies provide utilities using the HUI and 17D, these measures are less desirable than the mapped SF-36 for populating a NICE reference case analysis. In addition, the SF-36 studies on average measured outcomes later in life than the 17D or HUI studies, suggesting that the utilities obtained from these studies would be more relevant when extrapolated across the lifespan.

Three of these five studies are Norwegian.[121, 122, 126] One study restricts their population to ≤ 28 weeks gestational age, or ≤ 1000g birthweight.[121] This appears too limiting to capture the outcomes of all preterm children. The other two consider a more generous birthweight range of ≤ 1500g, but have small preterm sample sizes of 35 and 43 respectively.[122, 126]

Of the remaining two studies, one has SF-36 measures for a very low birthweight sample of 79 from the UK (assessed between ages 19 and 22).[120] The other uses a larger sample of 126 preterm children from New Zealand (assessed at age 31), whose mothers were participants in the Auckland Steroid Trial.[125] 21 of these individuals had RDS in infancy, which may lead to a small downward bias in the quality of life scores, though it is possible that this bias had been diminished by the time the participants were assessed.

In summary, the five studies that report SF-36 quality of life scores provide mapped EQ-5D utilities that can be used for preterm survivors. The studies that used the 17D or HUI measures were seen as less desirable, as they cannot be mapped in a straightforward way to EQ-5D utilities.

A4.9 Statistical analysis of the effects of birthweight on utility

In order to determine whether it was necessary to incorporate a utility reduction for lower birthweight, regression analysis was performed on a dataset obtained from one of the authors of a Canadian study.[116] The data contained utility, as measured by HUI3, assessed

at 3 life stages (adolescence, young adulthood, mature adulthood). Along with this, we were supplied with data on birthweight, sex, and gestational age. Data from 290 individuals was provided, although some had missing sex information, and many did not respond in all three life stages. Mean birthweight in the sample was 2047.9g (min = 560g, max = 4734g), and mean gestational age was 33.2 weeks (min = 23 weeks, max = 40 weeks).

A random effects GLS panel data estimator was used[4] to estimate the following general model:

$$u_{it} = \alpha + \beta B_i + \gamma X_i + \theta D_t + \nu_i + \epsilon_{it}$$

where u = utility, B = a vector containing birthweight and squared birthweight, X = a vector containing sex and gestational age, D = a vector of time dummies, v = an unobservable fixed effect, and ε = the idiosyncratic error term. Table 66 shows the estimates of five different model specifications.

Table 66 Random effects GLS estimates of utility on birthweight

	Dependent Variable: Utility (HUI3)				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Birthweight (g)	9.00e-05*	9.07e-05*	0.00011	0.00011	0.0000292
	(4.96e-05)	(4.96e-05)	(9.06e-05)	(9.05e-05)	(2.20e-05)
Birthweight squared	-1.08E-08	-1.11E-08	-1.4E-08	-1.4E-08	
	(1.10e-08)	(1.10e-08)	(1.52e-08)	(1.52e-08)	
Young adult		0.00196		0.00216	
		(0.0144)		(0.0145)	
Mature adult		-0.0531***		-0.0517***	-0.0528***
		(0.0167)		(0.0167)	(0.0149)
Gestational age (wks)			-0.0015	-0.0014	0.00246
			(0.00605)	(0.00605)	(0.00436)
Male			0.0463**	0.0440**	0.0446**
			(0.0198)	(0.0198)	(0.0198)
Constant	0.705***	0.716***	0.711***	0.721***	0.676***
	(0.0397)	(0.0403)	(0.116)	(0.116)	(0.105)
Observations	714	714	713	713	713
Number of individuals	287	287	286	286	286

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

[4] The 'between' estimator was also used, but results are not reported as they did not differ substantially from the GLS estimates, and because random effects GLS is a more efficient estimator in general.

The coefficient for birthweight squared was not significantly different from 0 in any of the five specifications. Specification (5) was performed as a result of removing the non-significant young adult dummy and the squared birthweight variable from (4), in order to compare model fit. A likelihood ratio test comparing these two models[5] resulted in a χ^2 statistic of 0.9 ($p = 0.6368$). Therefore, there is insufficient evidence for a quadratic relationship between birthweight and utility. According to (5), the marginal impact of a 150g reduction in birthweight is a utility reduction of 0.004.

The linear birthweight coefficient was only significantly different from 0 at the 10% level in specifications where sex and gestational age were not included. In order to test whether the simple specification (1) is equally valid to (4), another likelihood ratio test was performed. The test statistic of 18.16 ($p = 0.0011$) suggests we should reject that specification (1) is equally suitable to specification (4).

Based on this analysis, we conclude that there is insufficient evidence of a strong enough birthweight effect on utility to warrant inclusion in the economic model.

[5] These had to be re-estimated using maximum likelihood estimation (MLE) in order to obtain the log-likelihoods necessary to calculate the likelihood ratio test statistic. The parameter estimates were identical when using MLE and GLS.