National Institute for Health and Care Excellence

DIAGNOSTICS ASSESSMENT PROGRAMME

Evidence overview

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

This overview summarises the key issues for the diagnostics advisory committee's consideration. It is intended to be read with NICE's final scope for the assessment and the diagnostics assessment report. There is a glossary of terms in appendix B.

1 Background

1.1 Introduction

The purpose of this assessment is to evaluate the clinical and cost effectiveness of biomarker tests (phosphorylated IGFBP-1 using Actim Partus, placental alpha microglobuline-1 using PartoSure and quantitative fetal fibronectin using the Rapid fFN 10Q Cassette Kit) to help diagnose preterm labour in women with intact amniotic membranes. These tests may be used instead of fetal fibronectin testing (using a threshold of 50 nanograms/millilitre [ng/ml]) or clinical assessment alone, when transvaginal ultrasound measurement of cervical length is not available or acceptable. The results of the biomarker tests are designed to help clinicians decide whether women can be sent home or need to be admitted to hospital for treatment to delay birth and improve neonatal outcomes.

Issue date: February 2018

Provisional recommendations on using these technologies will be formulated by the diagnostics advisory committee at the committee meeting on 8 February 2018.

1.2 Scope of the assessment

Table 1 Scope of the assessment

Decision question	What is the clinical and cost effectiveness of tests used to help diagnose preterm labour in women with intact membranes?
Populations	Women with signs and symptoms of preterm labour and intact amniotic membranes who are 36^{+6} weeks pregnant or less, who are not in established labour, and for whom a transvaginal ultrasound is not available or acceptable. When data permit, subgroups based on gestational age will be considered.
Interventions	PartoSure with clinical assessment of symptoms. Actim Partus with clinical assessment of symptoms. Rapid fFN 10Q Cassette Kit, used with thresholds other than 50 ng/ml, with clinical assessment of symptoms.
Comparator	Fetal fibronectin, used with a threshold of 50 ng/ml, with clinical assessment of symptoms. Clinical assessment of symptoms alone.
Healthcare setting	Secondary care.
Outcomes	Intermediate measures to be considered may include: • diagnostic accuracy for preterm birth within 48 hours and 7 days • time to test result • test failure rate • number of women admitted to hospital • number of re-presentations to hospital within 48 hours and 7 days • number of women who have tocolytics or 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 appropriately (that is, they deliver within 7 days following treatment)

	• number of women treated inappropriately (that is, they do not deliver within 7 days following treatment) • effect on neonatal intensive care resource planning • gestational age at birth.
	Clinical outcomes to be considered may include: • perinatal mortality • neonatal morbidity and mortality • long-term health problems in the child • maternal morbidity and mortality.
	Patient-reported outcomes to be considered may include: • health-related quality of life • anxiety associated with confidence in the test results.
	Costs will be considered from an NHS and personal social services perspective. Costs to be considered may include: • cost of the test kit • cost of staff time to have training and do the testing • costs associated with resource use, such as hospital admissions and length of stay, treatments, and transfers between hospitals • costs associated with treating short-term neonatal health problems • costs associated with treating long-term effects of preterm birth.
	The cost effectiveness of interventions should be expressed in terms of incremental cost per quality-adjusted life year.
Time horizon	The time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared.

Further details including descriptions of the interventions, comparators, care pathway and outcomes can be found in the final scope.

2 The evidence

This section summarises data from the diagnostics assessment report (DAR) compiled by the external assessment group (EAG).

2.1 Clinical effectiveness

The EAG did a systematic review of the clinical effectiveness evidence for Actim Partus, PartoSure and quantitative fetal fibronectin using the Rapid fFN 10Q Cassette Kit. This included 2 reviews; 1 for diagnostic accuracy and 1 for clinical outcomes. Details of the systematic review can be found starting on page 40 of the DAR.

For the diagnostic accuracy review, studies were included if:

- they recruited women with signs and symptoms of preterm labour who were not in established labour and who had intact amniotic membranes
- the population was described as preterm
- twin or multiple pregnancies made up 20% or less of the total population recruited
- at least 1 index test was reported and at least 1 of the following reference standards or comparators was included:
  - preterm delivery within 48 hours
  - preterm delivery within 7 days
  - clinical assessment of symptoms
  - fetal fibronectin at a threshold of 50 ng/ml.
- they were prospective or retrospective diagnostic accuracy studies with random or consecutively recruited participants; both single- and two-gate[1] designs were eligible.

Full details of the inclusion and exclusion criteria can be found starting on page 42 of the DAR.

All studies included in the diagnostic accuracy review were appraised using the QUADAS-2 tool. In total, 20 studies met the inclusion criteria for the diagnostic accuracy review. The EAG also searched for studies in which

[1] A single-gate study recruits patients whose disease status is unknown before testing (a consecutive series) whereas a two-gate study recruits patients with the target condition and patients who do not have the target condition (a case-control study).

clinical outcomes were reported, but did not identify any studies. This search was restricted to controlled studies only because the EAG considered that uncontrolled study designs are likely to be susceptible to bias.

Study characteristics

Of the 20 diagnostic accuracy studies, data for more than 1 index test was reported in 2 studies (Hadzi-Lega et al. 2017, APOSTEL-1 2016), 16 studies assessed Actim Partus, 4 assessed PartoSure and 2 assessed fetal fibronectin at thresholds other than 50 ng/ml (APOSTEL-1, EUIFS 2016). All 20 studies assessed the index tests against a reference standard of preterm delivery within 7 days and 7 of these studies also assessed the index tests against a reference standard of preterm delivery within 48 hours.

QUADAS-2 assessment found that the risk of bias in all 20 studies was low, but some minor issues were identified:

- only 5 studies gave clear details of recruitment methods
- 2 studies used frozen samples, which is not in line with clinical practice (APOSTEL-1, Cooper 2012)
- 2 Actim Partus studies did not follow the company's instructions on sample collection (APOSTEL-1, Ting 2007)
- 8 studies did not provide details on blinding of clinicians and patients to the results of other tests that could influence the interpretation of the interventions.

Also, most of the studies did not provide details of whether women had tocolytics. Tocolytics are intended to delay preterm labour and may therefore have a direct effect on the reference standards (preterm delivery within 48 hours or 7 days). Full details of the critical appraisal of all 20 studies can be found starting on page 68 in the DAR.

As a result of differences in inclusion criteria among the studies, the characteristics of the study participants varied and this introduced heterogeneity:

- mean maternal age ranged from 25 to 31 years
- proportion of multiple gestations ranged from 0% to 20%
- mean number of previous term pregnancies ranged from 0.4 to 2.9 per person
- proportion of previous preterm deliveries ranged from 0% to 30%
- proportion of previous miscarriages ranged from 4% to 27%
- gestational age ranged from 20 weeks to 37 weeks.

Also, the prevalence of preterm birth in the studies ranged from 1.7% to 73.3% within 7 days and from 2.4% to 58.3% within 48 hours. These wide ranges are likely to reflect differences in the selection criteria for study participants. Further, the reporting of mode of delivery, that is, whether delivery was spontaneous or as a result of medical intervention, varied between studies. Only 11 studies provided details on mode of delivery; 4 excluded participants from test accuracy calculations if birth occurred because of medical intervention before the reference standard (7 days or 48 hours). Full details of participant characteristics in the included studies can be found in table 4 of the DAR.

Diagnostic accuracy

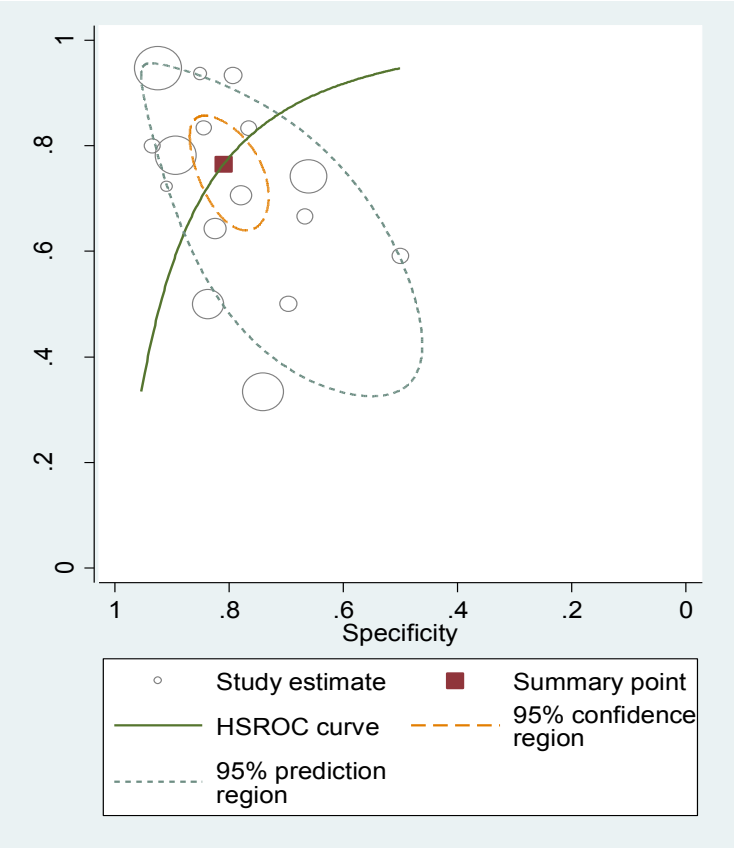
Delivery within 7 days: Actim Partus

There were 16 studies that reported data for Actim Partus. The prevalence of birth within 7 days of testing ranged from 1.7% to 73.3%. Across the studies sensitivity estimates ranged from 33.3% to 94.7%. The 3 studies (Cooper 2012, Danti 2011, Riboni 2011) with the lowest sensitivity estimates also had a lower prevalence of preterm birth (ranging from 1.7% to 6.7%) than the other studies (ranging from 9.8% to 73.3%). Specificity ranged from 50.0% to

93.5%. The EAG did not identify any major differences in methods or participant characteristics in the 3 studies with the lowest specificity estimates.

The EAG meta-analysed the results of these 16 studies using a random effects model. The pooled analysis estimated a sensitivity of 77% (95% confidence interval [CI] 68% to 83%) and a specificity of 81% (95% CI 76% to 85%). A receiver operating characteristic (ROC) plot summarising these results can be seen in figure 1.

Figure 1 ROC plot for Actim Partus against the 7-day reference standard



Key: HSROC, Hierarchical summary receiver-operating characteristic

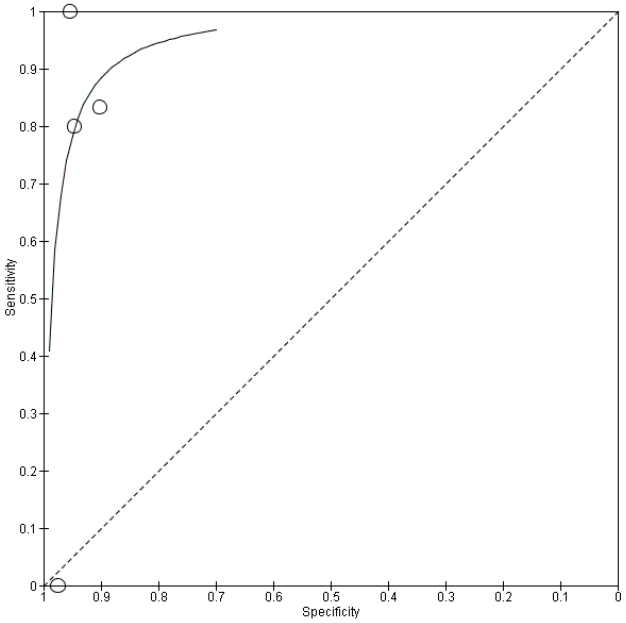
There were 6 studies that tested each sample with both Actim Partus and fetal fibronectin at a threshold of 50 ng/ml. Using delivery within 7 days as the reference standard, sensitivity for Actim Partus was lower than for fetal fibronectin in 1 study (APOSTEL-1 2016), higher in 2 studies (Ting 2017, Tripathi 2016) and the same for both tests in the remaining 3 studies (Cooper

2012, Eroglu 2007, Riboni 2011). Specificity was higher for Actim Partus than for fetal fibronectin in 4 of the 6 studies, and lower in the 2 remaining studies (Cooper, Tripathi).

Delivery within 7 days: PartoSure

There were 4 studies that reported diagnostic accuracy for PartoSure. The prevalence of birth within 7 days of testing ranged from 2.4% to 17.2%. Specificity was similar across studies, ranging from 90.2% to 97.5%, but sensitivity ranged from 0% to 100%. Werlen et al. (2015) reported 0% sensitivity because only 1 of 41 participants tested positive and this was a false positive. Meta-analysis of the 4 studies using a mixed effects multilevel logistic regression estimated a sensitivity of 83% (95% CI 61% to 94%) and a specificity of 95% (95% CI 89% to 98%). A ROC plot summarising these results can be seen in figure 2.

Figure 2 ROC plot for PartoSure against the 7-day reference standard



Nikolova et al. (2015) assessed fetal fibronectin at a threshold of 50 ng/ml and PartoSure in the same sample. Against the 7-day reference standard, sensitivity for PartoSure was 80% (95% CI 63.1 to 91.6) and for fetal fibronectin it was 50% (95% CI 21.1% to 78.95). Specificity for PartoSure was

94.6% (95% CI 90.1% to 97.5%) and for fetal fibronectin it was 72.2% (95% CI 58.4% to 83.5%).

Quantitative fetal fibronectin

There were 2 studies (APOSTEL-1 2016, EUIFS 2016) that reported diagnostic accuracy for quantitative fetal fibronectin. The prevalence of preterm birth within 7 days ranged from 10.5 % (EUIFS) to 19.7% (APOSTEL-1). In both studies sensitivity decreased as the threshold increased and specificity increased as the threshold increased. Results can be seen in table 2.

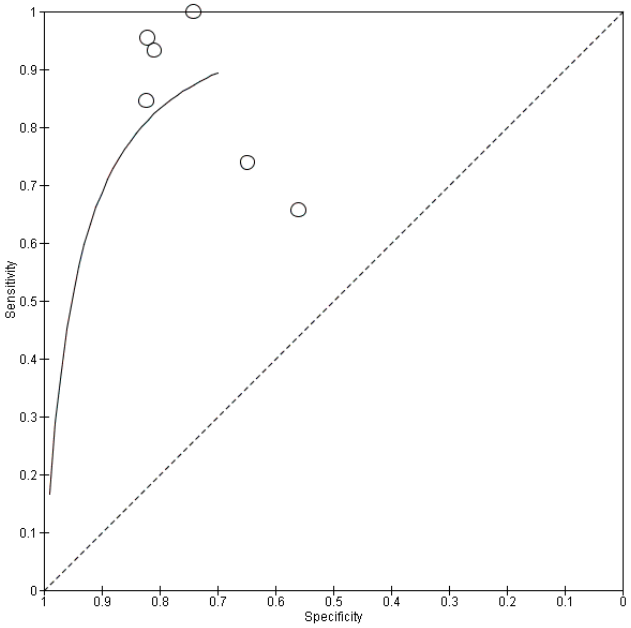
Table 2 Diagnostic accuracy of fetal fibronectin against the 7-day reference standard

Study	Threshold (ng/ml)	Sensitivity % (95% CI)	Specificity % (95% CI)
APOSTEL-1 (n=350)	10	95.7 (87.8, 99.1)	42.3 (36.5, 48.4)
	200	71.0 (58.8, 81.3)	83.6 (78.8, 87.8)
	500	42.0 (30.2, 54.5)	95.7 (92.7, 97.8)
EUIFS (n=455)	10	93.8 (82.8, 98.7)	32.2 (27.7, 37.0)
	200	70.8 (55.9, 83.0)	78.6 (74.3, 82.5)
	500	29.2 (17.0, 44.1)	94.3 (91.6, 96.4)
Abbreviations: CI, confidence interval; ng/ml, nanograms per millilitre			

Delivery within 48 hours: Actim Partus

There were 6 studies that assessed the diagnostic accuracy of Actim Partus. The prevalence of delivery within 48 hours of testing ranged from 5.3% to 58.3%. Sensitivity estimates ranged from 65.7% to 100.0% and specificity estimates ranged from 56.0% to 82.4%. The results were pooled using a mixed effects multilevel logistic regression. Pooled analyses of the 6 studies estimated a sensitivity of 87% (95% CI 74% to 94%) and a specificity of 73% (95% CI 62% to 82%). A ROC plot summarising these results can be seen in figure 3.

Figure 3 ROC plot for Actim Partus against the 48-hour delivery reference standard



Delivery within 48 hours: PartoSure

Only 1 study (Werlen et al. 2015) assessed the diagnostic accuracy of PartoSure against the 48-hour reference standard. The prevalence of preterm birth was 2.4%. Sensitivity was 0% (95% CI 0% to 97.5%) and specificity was 97.5% (95% CI 86.8% to 99.9%). Sensitivity was 0% because only 1 of 41 participants tested positive and this was a false positive.

Accuracy of comparator (fetal fibronectin with a threshold 50 ng/ml)

The EAG also did a non-systematic review of studies reporting diagnostic accuracy data for 1 of the comparators, fetal fibronectin at a threshold of 50 ng/ml. This review only included studies that were included in the diagnostic accuracy systematic review done for the intervention tests and that also reported data for this comparator. The EAG assessed the generalisability of these data by comparing them with results from 3 recently published systematic reviews of fetal fibronectin.

Of the 20 studies included in the diagnostic accuracy review, 8 reported accuracy data for quantitative fetal fibronectin at 50 ng/ml. Of these, 2 studies

(APOSTEL-1 2016, EUIFS 2016) used a quantitative fetal fibronectin test, 3 used the QuikCheck qualitative fetal fibronectin test (Eroglu 2007, Nikolova 2015, Tripathi 2016), 1 used an ELISA-based laboratory test (Riboni 2011), and in the remaining 2 studies (Cooper 2012, Ting 2007) it was not clear which fetal fibronectin test was used.

For the 8 studies looking at the diagnostic accuracy of fetal fibronectin at a threshold of 50 ng/ml against a 7-day reference standard, sensitivity ranged from 23.8% to 91.3% and specificity ranged from 62.2% to 99.1%.

The results for diagnostic accuracy of fetal fibronectin at a threshold of 50 ng/ml from the EAG's review were similar to those from the 3 existing literature reviews (see table 3).

Table 3 EAG diagnostic accuracy results for fetal fibronectin 50 ng/ml compared with existing literature reviews

	Sanchez-Ramos et al. (2009), 32 studies	Boots et al. (2014), 38 studies	NICE preterm labour and birth guideline 2015 20 studies	EAG identified studies
Sensitivity of fFN 50 ng/ml % (95% CI)	Pooled 76.1 (69.1 to 81.9)	Pooled 75 (69 to 80)	Ranged from 56 to 100 (95% CIs not reported)	Ranged from 23.8 (17.3 to 31.4) to 91.3 (82.0 to 96.7)
Specificity of fFN 50 ng/ml % (95% CI)	Pooled 81.9 (78.9 to 84.5)	Pooled 79 (76 to 83)	Ranged from 61.9 (59.6 to 62.5) to 92% (95% CI not reported)	Ranged from 62.2 (57.3 to 66.9) to 99.1 (97.3 to 99.8)
Abbreviations: CI, confidence interval; EAG, external assessment group; fFN, fetal fibronectin; ng/ml, nanograms per millilitre				

Diagnostic accuracy data informing the economic model

The EAG concluded that there was too much heterogeneity in the pooled results to use them for indirect comparisons between tests in the economic modelling. It decided to prioritise studies that reported results for more than 1 test in the same population. Two studies (APOSTEL-1 2016, Hadzi-Lega et al. 2017) identified in the systematic review were used in the base case for the economic model, because the EAG judged these to be the most suitable. Further details of the study characteristics are given in table 4.

Table 4 Study characteristics and results for studies used in the economic model base case

	APOSTEL-1 (Bruijn et al. 2016) n=350	Hadzi-Lega et al. (2017) n=57
Country	Netherlands, 10 centres	Macedonia, 1 centre
Mean maternal age (years)	29.9	27
Mean gestational age (weeks)	29	31
Proportion of multiple gestations (%)	20	0
Previous preterm delivery (%)	23	Not reported
Index test(s)	Actim Partus, quantitative fetal fibronectin	Actim Partus, PartoSure
Comparator	Fetal fibronectin 50 ng/ml	–
Actim Partus sensitivity within 7 days % (95% CI)	78.3 (66.7 to 87.3)	83.3 (35.9 to 99.6)
Actim Partus specificity within 7 days % (95% CI)	89.3 (85.1 to 92.7)	76.5 (62.5 to 87.2)
PartoSure sensitivity within 7 days % (95% CI)	n/a	83.3 (35.9 to 99.6)
PartoSure specificity within 7 days % (95% CI)	n/a	90.2 (78.6 to 96.7)
Diagnostic accuracy for fetal fibronectin at thresholds other than 50 ng/ml % (95% CI)	10 ng/ml: sensitivity 95.7 (87.8 to 99.1); specificity 42.3 (36.5 to 48.4) 200 ng/ml: sensitivity 71.0 (58.8 to 81.3); specificity 83.6 (78.8 to 87.8) 500 ng/ml: sensitivity 42.0 (30.2 to 54.5); specificity 95.7 (92.7 to 97.8)	n/a
Abbreviations: CI, confidence interval; n/a, not applicable; ng/ml, nanograms per millilitre		

2.2 Costs and cost effectiveness

The EAG did a search to identify existing studies investigating the cost effectiveness of Actim Partus, PartoSure, and quantitative fetal fibronectin. The EAG also constructed a de novo economic model to assess the cost effectiveness of the biomarker tests.

Systematic review of cost-effectiveness evidence

Cost and resource use studies

The EAG carried out a systematic review to identify previous studies of Actim Partus, PartoSure and fetal fibronectin that reported costs or resource use associated with the tests. The quality of the included studies was assessed using the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist. One study (Gibson et al. 2013) was identified from a conference abstract of an unpublished MSc report. The EAG obtained access to the full version of the dissertation for the review.

Gibson et al. assessed the effect of a fetal fibronectin test (threshold 50 ng/ml) on the use of antenatal corticosteroids and tocolysis in 306 women at St. Thomas' Hospital in London. It did not aim to assess the cost effectiveness of different testing strategies, but the EAG noted that it provided useful information for modelling use of steroids and fetal fibronectin at different thresholds. Compliance with the fetal fibronectin protocol was assessed and 67% of women who tested positive and 6% of women who tested negative had antenatal corticosteroids. Two (6%) of the women who tested positive delivered within 7 days and 10 (29%) before 37 weeks. The study also reported data on the use of tocolytics; 14% of women who tested positive and 2% of women who tested negative had tocolytics. Of those who tested positive and had tocolytics, 1 (10%) delivered within 7 days and 4 (40%) delivered before 37 weeks.

The EAG also identified 3 observational cost–minimisation studies that reported costs and resource use data, but these were published over 10 years

ago and it is not certain whether the protocols used in the studies reflect current clinical practice. Further details of these studies can be found starting on page 111 of the DAR.

Model-based studies

The EAG also carried out a search for existing economic models and identified 6 models. These varied by type of study, length of time horizon, obstetrician's compliance with test results and corresponding treatment, and the type of treatment modelled. Full details of the review can be found starting on page 113 of the DAR.

A cost–minimisation modelling approach was used in 2 studies (Chuck and Nguyen 2015, Deshpande et al. 2013). Chuck and Nguyen looked at the cost of adopting fetal fibronectin in Alberta, Canada and estimated that introducing the test between 2008 and 2013 led to 27 extra ambulance transfers, 1 less hospital admission, and 143 more hospital days for women not in labour. They concluded that introducing the test increased costs by US$4 million. The EAG noted that costs and benefits associated with fetal fibronectin testing from additional false negatives and true positives were not taken into account. The Deshpande et al. study was done in the UK. It modelled the cost difference between fetal fibronectin 50 ng/ml plus clinical examination compared with clinical examination alone. The study found that the rapid fetal fibronectin saved the NHS £23.88 per patient.

Cost-effectiveness modelling was used in 3 studies (Boyd et al. 2011, Mozurkewich et al. 2000 and van Baaren et al. 2017) and in NICE's guideline on preterm labour and birth. The results can be found on page 119 of the DAR. The most relevant model was the one used for NICE's guideline on preterm labour and birth (the NICE guideline model). The model was hypothetical and assessed what the specificity and sensitivity of the tests (cervical length measurement by ultrasound, Actim Partus and fetal fibronectin) would need to be for them to be considered cost effective compared with a no testing, treat-all strategy. It accounted for the effect of

gestational age on test accuracy and found that testing was not cost effective below 30 weeks' gestation. The key assumptions in the NICE guideline model were:

- the choice of diagnostic strategy had no significant effect on the mother's health outcomes
- clinicians did not deviate from the diagnostic protocol
- neonatal morbidity outcomes were based on respiratory distress syndrome and intraventricular haemorrhage
- the lifetime quality of life and costs are the same for both full-term and preterm infants.

The model results suggested that at lower gestational ages a treat-all strategy was cost effective, but as gestational age increased a diagnostic test was more likely to be cost effective.

Economic analysis

The EAG developed a de novo economic model to evaluate the cost effectiveness of Actim Partus, PartoSure and quantitative fetal fibronectin using the Rapid fFN 10Q Cassette Kit at thresholds other than 50 ng/ml compared with fetal fibronectin at 50 ng/ml. The EAG based their model on the NICE guideline for preterm labour and birth model. But it noted that several parameters used to populate the model, particularly utility and adverse event values, would need updating for this assessment. The EAG also noted that clinical opinion suggests the cost of neonatal intensive care (based on the British Association of Perinatal Medicine levels 1 to 4) used in the model was low, and that NHS tariffs for the 4 levels of care may be more accurate. Other areas that the EAG identified as important to expand on included:

- mortality benefits associated with reducing the rate of false negatives
- treatment costs associated with positive test results
- clinicians' compliance with treatment protocols.

Model structure

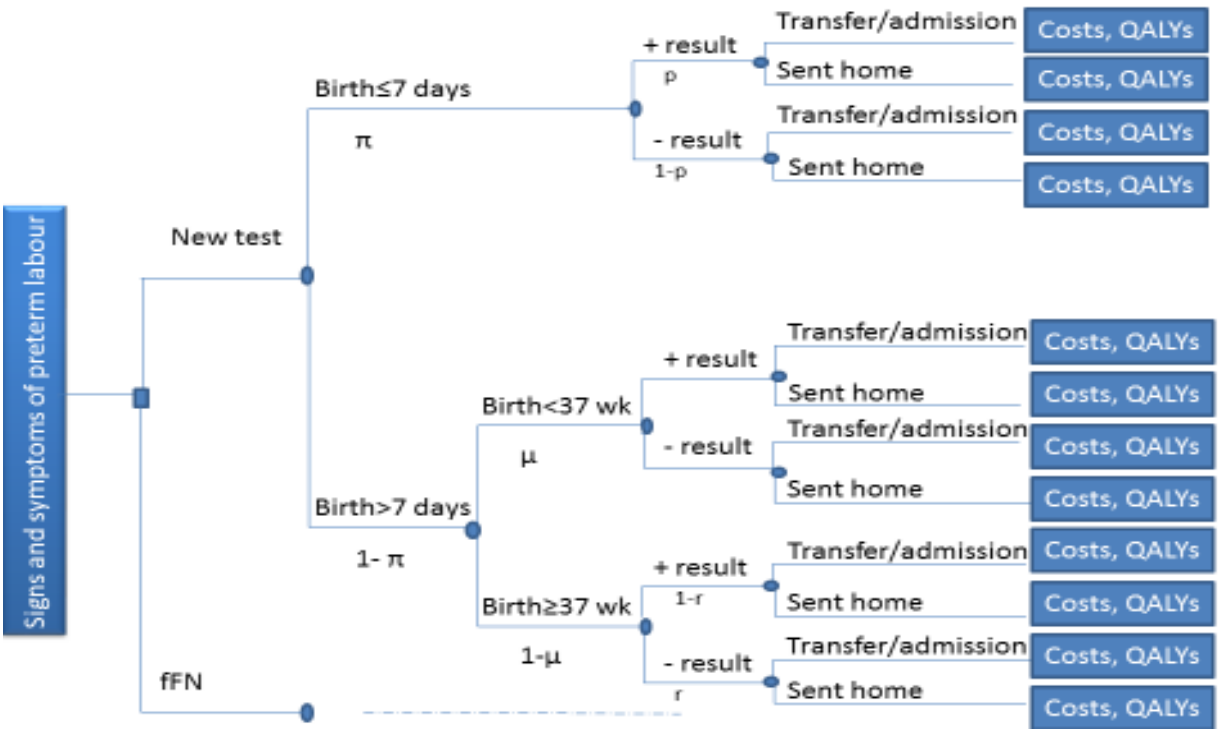
A decision tree structure similar to the NICE guideline model was used. It included a diagnostic phase followed by treatment and long-term outcomes. The model was based on the treatment pathway starting with an assessment of preterm labour, and models the decision of whether to offer corticosteroids or admit to hospital or discharge home, or a combination of these. It took the perspective of the NHS and personal social services and had a lifetime time horizon (100 years). A discount rate of 3.5% was applied to both costs and effects.

The population was women with intact membranes presenting with symptoms of threatened preterm labour who were between 24 and 36 weeks' gestation. It was assumed that, before entering the model, a clinical assessment had been done that could not rule out preterm labour. The decision tree had 2 main branches:

- one for the interventions (Actim Partus, PartoSure, and quantitative fetal fibronectin at thresholds of 10, 200 and 500 ng/ml) or for the no-test treat-all strategy, which assumes that all women entering the model are admitted to hospital
- one for the comparator (fetal fibronectin at a threshold of 50 ng/ml).

The structure of the decision tree is shown in figure 4. Longer-term costs and QALYs were calculated for each branch of the decision tree.

Figure 4 Decision tree structure



Model inputs

The model was populated with data from the diagnostic accuracy review, published literature and expert opinion. Full details of the model inputs can be seen starting on page 138 of the DAR.

Accuracy data

Accuracy data used in the economic model base case was taken from 2 main studies, APOSTEL-1. Direct estimates of diagnostic accuracy were obtained for fetal fibronectin and Actim Partus from APOSTEL-1. None of the included studies directly compared PartoSure with fetal fibronectin so this was estimated using data from APOSTEL-1 and Hadzi-Lega et al. Sensitivity analyses were also done using data from alternative sources: Cooper et al. 2012, Abbott et al. 2013 and an EAG meta-analysis. Full details of the estimates can be found in table 5.

Table 5 Diagnostic accuracy values used in the economic model

Study	Diagnostic test (threshold)	Sensitivity %	Specificity %
Bruijn et al. 2016 (APOSTEL-1), n=350 (base case)	fFN (10 ng/ml)	0.957	0.423
	fFN (50 ng/ml)	0.913	0.648
	fFN (200 ng/ml)	0.710	0.836
	fFN (500 ng/ml)	0.420	0.957
	Actim Partus	0.783	0.893
Hadzi-Lega et al. 2017 n=57 (base case)	PartoSure	0.833	0.902
	Actim Partus	0.833	0.765
Cooper et al. 2012 n=349 (scenario)	Actim Partus	0.333	0.741
	fFN (50 ng/ml)	0.333	0.898
Abbott et al. 2013 n=299 (scenario)	fFN (10 ng/ml)	0.778	0.576
	fFN (50 ng/ml)	0.778	0.790
	fFN (200 ng/ml)	0.778	0.931
	fFN (500 ng/ml)	0.556	0.972
EAG meta-analysis n=963 (scenario)	Actim Partus	0.832	0.879
	fFN (50 ng/ml)	0.683	0.872
Abbreviations: fFN, fetal fibronectin; ng/ml, nanograms per millilitre			

Costs

Costs included in the model can be seen in table 6. All costs were inflated to 2016 prices using the healthcare prices index in the Personal Social Services Research Unit (PSSRU) publication of unit costs of health and social care.

The EAG adjusted costs from the NICE guideline model to obtain an estimate for the long-term health costs of intraventricular haemorrhage. The NICE guideline model estimated a cost of £23,700, but the EAG considered that this could be underestimated and proposed that the cost is at least £79,000, but could be as high as £114,648.

Table 6 Model cost parameter values

Cost parameter	Unit cost (£)	Definition	Source
Fetal fibronectin test	65	15 minutes of midwife time, cost of test £35.	PSSRU Hologic 2017 request for information
pIGFBP-1	35	10 minutes of midwife time, cost of test £15.	PSSRU Alere 2017 request for information
PAMG-1	52	10 minutes of midwife time, cost of test £32.	PSSRU Parsagen 2017 request for information
Maternal steroid injection	5		UCLH 2012; Parisaei 2016
Tocolytics: atosiban plus infusion equipment	362	Syringe pump, syringe and giving set, cost of solution for infusion atosiban acetate 37.5 mg/5 ml concentrate for solution for infusion vials, 1 vial £52.82.	Parisaei et al. 2016 BNF 2016
Inpatient hospital	1,325	Median length of hospital stay 2 days.	Parisaei 2016
In utero transfer	965	London Ambulance Service, 6 hours of a matron's time to arrange transfer, £372 (Curtis and Burns, 2016).	Parisaei 2016 Gale et al. 2012
Long-term healthcare costs of IVH	114,648	Downstream healthcare costs.	Updated from NICE guideline 2015
Neonatal hospital costs of preterm survivors discharged home or admitted to ward	32,435	Costs of BAPM levels 1, 2, 3, 4, 5 at 2014/15 NHS tariffs. Mean overall length of stay of 46 days.	EAG analysis of data from Badger neonatal outcome national dataset
Neonatal hospital costs RDS	5,587	<u>Base case:</u> OLS-adjusted difference in neonatal hospital costs between infants with and without days spent in BAPM level of care 1. <u>Scenario:</u> 1. Downstream healthcare costs; NHS reference costs 2011/12, XB01Z Paediatric	<u>Base case:</u> Badger 2014/2015 National tariffs for BAPM levels 1, 2, 3, 4, 5 2014/15 <u>Scenario</u>: 1. NICE guideline 2015

		critical care, intensive care, ECMO/ECLS. 2. Hospitalisation cost, 2008 Canadian dollars, cost of medical and pharmaceutical services also given.	
Additional neonatal hospital costs: infant dies before discharge	−22,834		Base case: calculated by EAG from Badger data Scenario: Khan et al. 2015
Abbreviations: BAPM, British Association of Perinatal Medicine; EAG, external assessment group; BNF, british national formulary; pIGFBP-1, phosphorylated insulin-like growth factor-binding protein 1; ECLS, extracorporeal life support; ECMO, extracorporeal membrane oxygenation; IVH, intraventricular haemorrhage; PAMG-1, placental alpha microglobulin-1; RDS, respiratory distress syndrome; PSSRU, personal social services research unit; OLS, ordinary least squares; UCLH, University College London Hospitals			

Health-related quality of life and QALY decrements

Health-related quality of life estimates were included for neonates in the base case, and a scenario analysis also included maternal health-related quality of life estimates. A systematic search identified 28 relevant papers, of which 4 were chosen to populate the economic model. A utility for severe persistent asthma was applied to 56% of children with respiratory distress syndrome based on clinical expert opinion. Full details of the utilities in the economic model can be seen in table 7.

Table 7 Utilities used in the economic model

Variable	Patient	Source	Measure	Utility	Range
'Severe' RDS (severe persistent asthma used as proxy)	Child	Carroll and Downs 2009	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. 2012	TTO	0.76	0.66, 0.84
Death	Child	Upper bound from Vandenbussche et al. 1999	SG	0	0, 0.02
Preterm survivor	Child	Cooke 2004	SF-36	0.879	0.846, 0.901
Mother with previous adverse child outcome	Mother	Couto et al. 2009	SF-36	0.644	0.556, 0.652
Mother with no adverse child outcome	Mother	Couto et al. 2009	SF-36	0.834	0.768, 0.843
Abbreviations: IVH, intraventricular haemorrhage; RDS, respiratory distress syndrome; SG, standard gamble; TTO time trade off					

Base-case results

For decision-making, the incremental cost-effectiveness ratios (ICERs) per QALY gained or lost will be considered. The following assumptions were applied in the base-case analysis:

- Clinical examination was not included as a comparator, because the population entered the model after a clinical examination which did not rule out preterm labour.
- All treatment decisions were driven by the test result, and clinical judgement did not override this.
- Diagnostic accuracy was equivalent across all gestational ages modelled.
- The prevalence of preterm birth within 7 days of testing was 3%, and the preterm birth rate, that is birth before 37 weeks, was 12.1%.

- Antenatal corticosteroids were only effective within 7 days of delivery, infants born more than 7 days after treatment did not benefit.
- In utero transfers were only available for women presenting to a level 1 or 2 hospital at less than 28 weeks' gestation.
- Tocolysis was used for all in utero transfers at gestational ages less than 28 weeks.
- Only intraventricular haemorrhage resulted in longer-term costs.
- Neonates who survived beyond 1 year had a long-term quality of life equivalent to the general population average.

The base-case results were given for groups of women presenting at the following mean gestational ages:

- 33 weeks
- 30 weeks
- 26 weeks.

Results were also stratified according to the level of neonatal care available at the place of birth. For example a level 1 unit would be a special care baby unit for neonates who need additional observations or help with feeding, whereas a level 3 unit is capable of providing full life support including ventilation (a neonatal intensive care unit). Regional neonatal centres may provide more than 1 level of care, but a district general hospital may only have a level 1 unit.

For women presenting at 30 weeks' gestation at a level 2 hospital most of the tests were cheaper and less effective than fetal fibronectin 50 ng/ml, apart from a treat-all strategy and fetal fibronectin at 10 ng/ml, which resulted in very small QALY gains and additional cost. PartoSure appeared to be the most cost-effective intervention compared with fetal fibronectin at a threshold of 50 ng/ml. The results can be seen in table 8 and the fully incremental analyses can be found in the DAR on page 159 table 33. Note that some of the index tests are cheaper and less effective than the comparator(s). An increase in these ICERs indicates an increase in savings per QALY lost.

Table 8 Base-case cost-effectiveness results 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**	**Inc. costs (£)**	**Inc. QALYs**	**ICER (£ per QALY)**	**Inc. costs (£)**	**Inc. QALYs**	**ICER (£ per QALY)**
Actim Partus	5,055	22.010	−1,116	−0.010	108,323*	−346	−0.006	56,033*
PartoSure	4,895	22.010	−1,276	−0.008	123,858	−506	−0.006	81,925*
Treat all	6,171	22.020	0	0	−	770	0.004	186,757
fFN 10 ng/ml	5,690	22.018	−481	−0.002	233,245*	289	0.002	140,270
fFN 50 ng/ml	5,690	22.016	−770	−0.004	186,757*	–	–	–
fFN 200 ng/ml	5,159	22.006	−1,012	−0.014	73,676*	−242	−0.010	25,213*
fFN 500 ng/ml	5,004	21.992	−1,167	−0.027	42,474*	−398	−0.023	17,013*

Abbreviations: ICER, incremental cost-effectiveness ratio; fFN, fetal fibronectin; QALY, quality-adjusted life years; Inc, incremental; ng/ml, nanograms per millilitre

* ICER is in the south-west quadrant in cost-effectiveness plane (cost saved per QALY lost)

The ICERs for women in a level 2 hospital at 26 weeks' gestation were lower than at 30 weeks. This was the case for ICERs that are in the south-west quadrant (which became less cost effective) and for those in the north-east quadrant (which became more cost effective). Full results of this analysis can be seen in table 35 page 161 of the DAR.

The ICERs for women in a level 2 hospital at 33 weeks' gestation were higher than at 30 weeks. This was the case for ICERs that are in the south-west quadrant (which became more cost effective) and for those in the north-east quadrant (which became less cost effective). Full results can be seen in table 38 page 163 of the DAR.

Analysis of alternative scenarios

The EAG explored the following scenarios:

- alternative sources of test accuracy data
- women presenting at hospitals with a level 3 unit
- only including neonatal phase costs
- only including diagnostic phase (until delivery) costs
- antenatal steroids having benefits when given more than 7 days before preterm delivery
- excluding hospital costs of infant death
- including the negative effect of a neonatal death on a woman's quality of life.

Alternative diagnostic accuracy data from Cooper et al. (2016) were used to calculate ICERs for Actim Partus compared with fetal fibronectin at a threshold of 50 ng/ml. This study included 349 women with a mean gestational age of 29^{+6} weeks from 2 centres in Canada. In this analysis Actim Partus was dominated by fetal fibronectin (that is, fetal fibronectin was more effective and less expensive). It was not clear which version of the fetal fibronectin test was used in this study. The EAG assumed that the ELISA version was used, which is no longer used in clinical practice. Also, alternative diagnostic accuracy data were obtained for fetal fibronectin at thresholds of 10, 200 and 500 ng/ml compared with fetal fibronectin at a threshold of 50 ng/ml from an unpublished study by Abbott et al. In this analysis a threshold of 10 ng/ml was [redacted] 50 ng/ml, 100 ng/ml [redacted] 50 ng/ml and 500 ng/ml had an ICER of [redacted] compared with 50 ng/ml.

When the time horizon was limited to the diagnostic phase only, that is excluding any postpartum outcomes, the analysis became a cost–minimisation analysis. PartoSure was the least costly option with £507 saved

per woman, followed by fetal fibronectin at a threshold of 500 ng/ml with a saving of £399, and Actim Partus with a saving of £347.

The scenario analysis with the greatest effect on the ICER was limiting the time horizon of the analysis to the first year after birth. This increased the ICER for all interventions by more than 20 times the base-case value. The greatest increase was for the treat-all strategy; it went from £186,757 per QALY gained in the base case to £4,930,444 per QALY gained (see table 9).

Allowing antenatal steroids to have benefits if given more than 7 days before birth also had a considerable effect on all the ICERs. This lowered the ICER for PartoSure, treat-all and fetal fibronectin at a threshold of 10 ng/ml. However it increased the ICER for Actim Partus and fetal fibronectin at a threshold of 200 ng/ml and 500 ng/ml (see table 9). None of the other scenarios significantly affected the results and further details can be found starting on page 175 of the DAR.

Table 9 ICERs compared with fetal fibronectin (threshold 50 ng/ml) for women presenting at 30 weeks' gestation (level 2 hospital)

Option	Base case (£)	Limiting the analysis to first year after birth (£)	ANS earlier than 7 days before preterm delivery has partial benefits (£)
Treat all	186,757	4,930,444	41,625
fFN 10 ng/ml	140,270	3,704,229	24,420
fFN 200 ng/ml	25,213*	669,308*	9,729*
fFN 500 ng/ml	17,013*	453, 004*	7,422*
Actim Partus	56,033*	1,482,263*	16,663*
PartoSure	81,925*	2,165,244*	128,511*

Abbreviations: ANS, antenatal corticosteroids; fFN, fetal fibronectin; ng/ml, nanogrammes per millilitre

* ICER represents the south-west quadrant in the cost-effectiveness plane (that is, a reduction in both costs and QALYs)

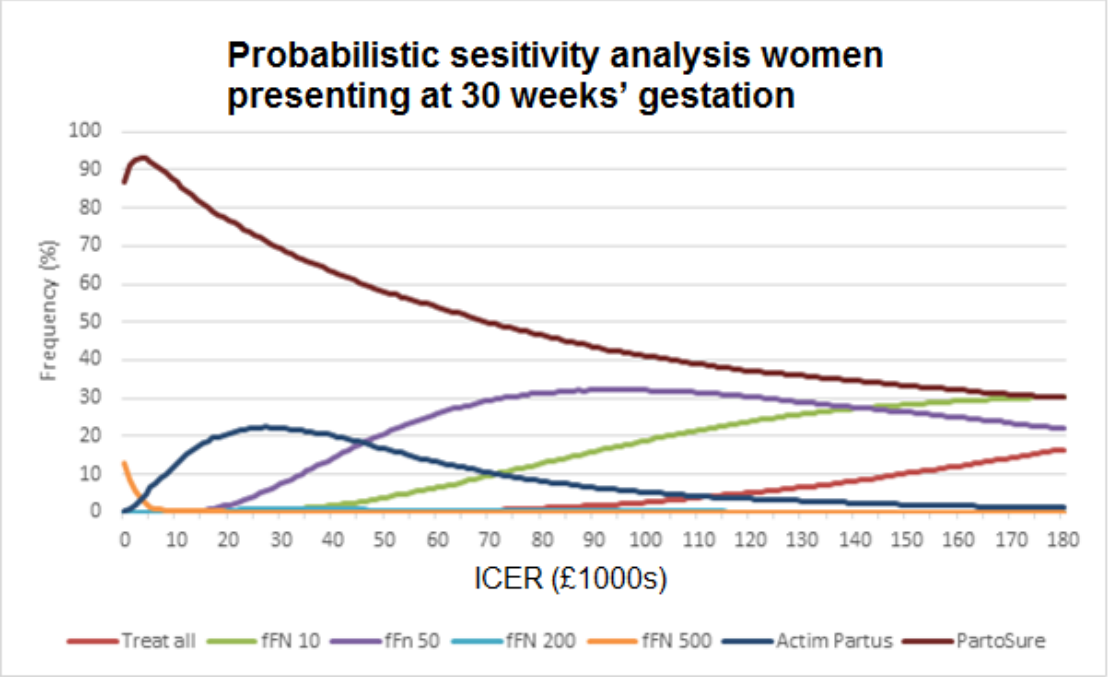
Sensitivity analyses

Deterministic sensitivity analyses were carried out by varying the base-case parameters by 20%. These analyses found that the ICERs were most sensitive to changes in health-related quality of life of preterm survivors. Other parameters affecting the ICERs were; cost of hospital admission, prevalence of preterm birth within 7 days, effectiveness of steroid treatment and baseline mortality risk. The results of the deterministic sensitivity analyses can be found starting on page 164 of the DAR.

Probabilistic sensitivity analysis was also carried out and the sampling distributions used for each parameter can be seen starting on page 40 of the DAR, in tables 22, 23 and 25. For women presenting at 30 weeks' gestation, at an ICER of £20,000 per QALY gained, PartoSure had a 76% probability of being the most cost-effective option, Actim Partus had a probability of 21% and a treat-all without testing strategy had a probability of 0%. The results of

this probabilistic sensitivity analysis can be seen in figure 5. PartoSure also had the greatest probability of being cost effective at an ICER of £20,000 per QALY gained for women presenting at 33 and 26 weeks' gestation.

Figure 5 Probabilistic analysis: women presenting at 30 weeks' gestation



3 Summary

3.1 Clinical effectiveness

There were 20 studies that provided diagnostic accuracy data for the interventions. The external assessment group (EAG) identified substantial heterogeneity between the studies, particularly in terms of participant characteristics and methodology of the studies. The prevalence of preterm birth varied widely, it is uncertain whether this affects the generalisability of the test accuracy results to the NHS. When possible, the EAG meta-analysed sensitivity and specificity estimates but because of the substantial heterogeneity the results of these analyses were highly uncertain. Because of this the EAG preferred to use test accuracy estimates from individual studies, which reported results for more than 1 index test or comparator. Of the

2 studies (APOSTEL-1 2016 and Hadzi-Lega et al. 2017) that looked at more than 1 index test, there were no substantial differences in test accuracy between PartoSure and Actim Partus in 1 study, although PartoSure had higher specificity. The other found that the relative accuracy of fetal fibronectin and Actim Partus varied, depending on what threshold was used for the fetal fibronectin test. The EAG was unable to conclude whether test accuracy differs based on gestational age within studies because results were not reported by these subgroups.

The EAG highlighted that there is a need for more studies assessing more than 1 index test in the same trial, which allow for more robust diagnostic accuracy data comparing the tests.

3.2 Cost effectiveness

The results of the economic analyses should be interpreted with caution, because of the wide differences in diagnostic accuracy for all of the index tests identified in the EAG's systematic review of clinical data.

The base-case results of the EAG's economic model for women presenting at 30 weeks' gestation showed that Actim Partus was cheaper and less effective than fetal fibronectin at a threshold of 50 ng/ml with an ICER of £56,033 saved per QALY lost, that is, a QALY loss of 0.006 and a cost saving of £346. PartoSure was cheaper than fetal fibronectin, with a cost saving of £160. It was as effective as Actim Partus, based on data from a small study (Hadzi-Lega et al. 2017). As the threshold chosen for fetal fibronectin decreases, its effectiveness increases and its cost increases. Compared with a threshold of 50 ng/ml, the ICERs for fetal fibronectin are £140,270 per QALY gained at a threshold of 10 ng/ml and £17,013 lost per QALY gained at a threshold of 500 ng/ml. The test most likely to be cost effective at an ICER of £20,000 per QALY gained was PartoSure, although all estimates of cost effectiveness for this test were based on indirect comparisons. The EAG found that the range of ICERs was wide in sensitivity and scenario analyses. The scenario

analyses that had the greatest effect on the results were limiting the model time horizon to the first year after birth and allowing the benefit of antenatal steroids to continue after 7 days from administration. The ICERs were most sensitive to changes in quality-of-life estimates applied to neonates who are born preterm and survive.

4 Issues for consideration

4.1 Clinical effectiveness

The prevalence of preterm birth ranged from 1.7% to 73.3% in the 20 studies included in the review. This is likely to affect the generalisability of the test accuracy results to the NHS.

The mode of delivery was not well reported in some studies. To assess the accuracy of the test against the reference standards (delivery within 48 hours or 7 days) it is important to be able to distinguish between deliveries that occurred spontaneously and those that occurred because of medical intervention. When medical intervention occurred within the reference standard timeframe, it was not possible to determine whether a spontaneous birth would have occurred, so it would be preferable to exclude these deliveries from the analysis. Therefore there is uncertainty around applying the reference standard when mode of delivery is not reported, and consequently the diagnostic accuracy estimates provided by the studies.

Treatment protocols varied between the included studies, which may have affected the reference standard. For example, treatment decisions in APOSTEL-1 2016 were based on the results of cervical length measurement and quantitative fetal fibronectin at a threshold of 50 ng/ml combined, and were not guided by the results of the interventions. In Hadzi-Lega et al. (2017) treatment decisions were based on standard of care at the hospital.

The EAG was unable to conclude whether test accuracy differs based on gestational age within studies because results were not reported by these

subgroups. The gestational age of women recruited to the studies varied; 14 out of 20 studies recruited women from as early as 24 weeks. Of the remaining 6 studies 4 recruited women as early as 20 weeks, and 2 recruited from 28 weeks.

Test accuracy data for fetal fibronectin using a 50 ng/ml threshold were identified from studies included in the diagnostic accuracy review of the interventions, and compared with recently published systematic reviews. This suggested that the EAG's results are broadly representative, but additional uncertainties may have been introduced by restricting the review to studies included in the diagnostic accuracy review.

A systematic review of end-to-end studies was carried out to identify evidence of the effect of the tests on clinical outcomes. This search was limited to controlled studies only because the EAG considered that uncontrolled study designs might be too prone to bias. It is therefore not certain whether relevant outcomes have been missed in the uncontrolled studies that were excluded. For example, implementation studies that report the effect of test results on clinical decision-making may have provided useful supplementary information.

4.2 Cost effectiveness

The EAG prioritised studies that reported data for more than 1 test for inclusion in the base case of the economic model, because of heterogeneity in the pooled estimates. The base case included 2 studies (APOSTEL-1 and Hadzi-Lega et al. 2017). APOSTEL-1 included 350 women with a mean gestational age of 29 weeks from 10 centres across the Netherlands. The second study (Hadzi-Lega et al.) included 57 women with a mean gestational age of 31 weeks from 1 centre in Macedonia. It is therefore uncertain if these studies included a population who would be broadly representative of NHS patients.

Diagnostic accuracy results for PartoSure were calculated indirectly based on data from 2 studies (Hadzi-Lega et al. and APOSTEL-1). No other sources of accuracy data for PartoSure were explored in the scenario analyses.

The model assumes that clinicians do not override test results when deciding whether to admit a woman for further treatment or to discharge them. Therefore, in the model all women with false negative results do not have antenatal steroids and do not have any treatment benefits.

The longer-term healthcare costs of intraventricular haemorrhage are high (£114,648) and these were estimated using a national dataset (Badger). The figure was obtained from an adjusted estimate of the cost difference between neonates who did and did not have neonatal intensive care. It is not certain if this estimate reflects the true cost of treating the effects of intraventricular haemorrhage in the longer-term.

Only limited data on the quality-of-life outcomes for women who have had a preterm delivery were available. This has been explored in the modelling, but it is not certain whether the true effect has been captured.

As noted in the issues for consideration on clinical effectiveness, the EAG's exclusion of clinical outcome studies that were uncontrolled, because of the risk of bias, may have prevented potentially useful real world data informing the economic modelling parameters and inputs.

5 Equality considerations

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others.

Pregnancy and maternity are protected characteristics under the Equality Act 2010. Children born before 37 weeks' gestation are at increased risk of

developing disabilities, which would lead to them being protected under the Act.

6 Implementation

The NICE Adoption team collated information from healthcare professionals working within NHS organisations who have experience of using biomarker tests to help diagnose preterm labour. Adoption barriers noted by experts were related to internal governance and procurement processes, which were reported to be challenging and time consuming.

7 Authors

Jessica Maloney
Topic Lead

Rebecca Albrow
Technical Adviser

February 2018

Appendix A: Sources of evidence considered in the preparation of the overview

A. The diagnostics assessment report for this assessment was prepared by Peninsula Technology Assessment Group (PenTAG):

- Varley-Campbell J, Mújica-Mota R, Coelho H et al. Biomarker tests to help diagnose preterm labour in women with intact membranes. 2018

B. The following organisations accepted the invitation to participate in this assessment as stakeholders. They were invited to attend the scoping workshop and to comment on the diagnostics assessment report.

Manufacturer(s) of technologies included in the final scope:

- Advanced Global Health
- Alere
- Medix Biochemica

Other commercial organisations:

- None

Professional groups and patient/carer groups:

- British Maternal and Fetal Medicine Society
- Royal College of Obstetricians and Gynaecologists
- National Childbirth Trust
- Tommy's

Research groups:

- None

Associated guideline groups:

- None

Others:

- Department of Health
- Healthcare Improvement Scotland
- Medicines and Healthcare products Regulatory Agency
- Medway NHS Foundation Trust
- NHS England
- Welsh Government

Appendix B: Glossary of terms

Diagnosed preterm labour

A woman is in diagnosed preterm labour if she is in suspected preterm labour and has had a positive diagnostic test for preterm labour.

Established preterm labour

A woman is in established preterm labour if she has progressive cervical dilatation from 4 cm with regular contractions.

Intraventricular haemorrhage

Bleeding into the fluid-filled areas (ventricles) inside the brain.

Suspected preterm labour

A woman is in suspected preterm labour if she has reported symptoms of preterm labour and has had a clinical assessment (including a speculum or digital vaginal examination) that confirms the possibility of preterm labour but rules out established labour.