

## Three biomarker tests to help diagnose preterm labour: a systematic review and economic evaluation

*Jo Varley-Campbell, Rubén Mújica-Mota, Helen Coelho, Neel Ocean, Max Barnish, David Packman, Sophie Dodman, Chris Cooper, Tristan Snowsill, Tracey Kay, Neil Liversedge, Michelle Parr, Lisa Knight, Chris Hyde, Andrew Shennan and Martin Hoyle on behalf of the Peninsula Technology Assessment Group (PenTAG)*



**National Institute for  
Health Research**



# Three biomarker tests to help diagnose preterm labour: a systematic review and economic evaluation

Jo Varley-Campbell,<sup>1\*</sup> Rubén Mújica-Mota,<sup>1</sup>  
Helen Coelho,<sup>1</sup> Neel Ocean,<sup>1</sup> Max Barnish,<sup>1</sup>  
David Packman,<sup>1</sup> Sophie Dodman,<sup>1</sup> Chris Cooper,<sup>1</sup>  
Tristan Snowsill,<sup>1,2</sup> Tracey Kay,<sup>3</sup> Neil Liversedge,<sup>3</sup>  
Michelle Parr,<sup>4</sup> Lisa Knight,<sup>3</sup> Chris Hyde,<sup>1</sup>  
Andrew Shennan<sup>5,6</sup> and Martin Hoyle<sup>1</sup> on behalf of  
the Peninsula Technology Assessment Group (PenTAG)

<sup>1</sup>Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, University of Exeter, Exeter, UK

<sup>2</sup>Health Economics Group, University of Exeter Medical School, University of Exeter, Exeter, UK

<sup>3</sup>Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

<sup>4</sup>Central Manchester University Hospital NHS Foundation Trust, Manchester, UK

<sup>5</sup>Department of Women and Children's Health, King's College London, London, UK

<sup>6</sup>Guy's and St Thomas' Hospital, London, UK

\*Corresponding author

**Declared competing interests of authors:** Andrew Shennan is an investigator in a number of trials/ studies related to preterm birth (the GlaxoSmithKline-funded NEWBORN tocolytic trial, the National Institute for Health Research-funded PETRA and QUIDS prediction studies, the Guy's and St Thomas' charity-funded EQUIPPT, the preterm management study and Tommy's charity-funded preterm birth studies). These studies include comparing PartoSure™ (Parsagen Diagnostics Inc., Boston, MA, USA) and the quantitative Fetal Fibronectin (fFN) Test (Hologic, Inc., Marlborough, MA, USA) and have been supported by free PartoSure samples from QUIAGEN and received financial support from Hologic, Inc. (fFN), paid to his institution to cover expenses of this comparison only. He has given lectures to internal staff at BioMedica [Actim® Partus (Medix Biochemica, Espoo, Finland)] and Hologic, Inc. (fFN), in the last 5 years and received financial support to cover expenses only for this when travelling to the USA and Finland.

Published March 2019

DOI: 10.3310/hta23130



This report should be referenced as follows:

Varley-Campbell J, Mújica-Mota R, Coelho H, Ocean N, Barnish M, Packman D, *et al.* Three biomarker tests to help diagnose preterm labour: a systematic review and economic evaluation. *Health Technol Assess* 2019;**23**(13).

*Health Technology Assessment* is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.



ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.513

*Health Technology Assessment* is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) ([www.publicationethics.org/](http://www.publicationethics.org/)).

Editorial contact: [journals.library@nhr.ac.uk](mailto:journals.library@nhr.ac.uk)

The full HTA archive is freely available to view online at [www.journalslibrary.nhr.ac.uk/hta](http://www.journalslibrary.nhr.ac.uk/hta). Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: [www.journalslibrary.nhr.ac.uk](http://www.journalslibrary.nhr.ac.uk)

## Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

## HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: <http://www.nets.nhr.ac.uk/programmes/hta>

## This report

The research reported in this issue of the journal was commissioned and funded by the HTA programme on behalf of NICE as project number 17/10/01. The protocol was agreed in July 2017. The assessment report began editorial review in January 2018 and was accepted for publication in July 2018. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Varley-Campbell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library ([www.journalslibrary.nhr.ac.uk](http://www.journalslibrary.nhr.ac.uk)), produced by Prepress Projects Ltd, Perth, Scotland ([www.prepress-projects.co.uk](http://www.prepress-projects.co.uk)).

## NIHR Journals Library Editor-in-Chief

**Professor Ken Stein** Chair of HTA and EME Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

## NIHR Journals Library Editors

**Professor Ken Stein** Chair of HTA and EME Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

**Professor Andrée Le May** Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals)

**Professor Matthias Beck** Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

**Dr Tessa Crilly** Director, Crystal Blue Consulting Ltd, UK

**Dr Eugenia Cronin** Senior Scientific Advisor, Wessex Institute, UK

**Dr Peter Davidson** Consultant Advisor, Wessex Institute, University of Southampton, UK

**Ms Tara Lamont** Scientific Advisor, NETSCC, UK

**Dr Catriona McDaid** Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

**Professor William McGuire** Professor of Child Health, Hull York Medical School, University of York, UK

**Professor Geoffrey Meads** Professor of Wellbeing Research, University of Winchester, UK

**Professor John Norrie** Chair in Medical Statistics, University of Edinburgh, UK

**Professor John Powell** Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

**Professor James Raftery** Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

**Dr Rob Riemsma** Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

**Professor Helen Roberts** Professor of Child Health Research, UCL Great Ormond Street Institute of Child Health, UK

**Professor Jonathan Ross** Professor of Sexual Health and HIV, University Hospital Birmingham, UK

**Professor Helen Snooks** Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

**Professor Jim Thornton** Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

**Professor Martin Underwood** Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of editors: [www.journalslibrary.nihr.ac.uk/about/editors](http://www.journalslibrary.nihr.ac.uk/about/editors)

**Editorial contact:** [journals.library@nihr.ac.uk](mailto:journals.library@nihr.ac.uk)



# Abstract

## Three biomarker tests to help diagnose preterm labour: a systematic review and economic evaluation

Jo Varley-Campbell,<sup>1\*</sup> Rubén Mújica-Mota,<sup>1</sup> Helen Coelho,<sup>1</sup> Neel Ocean,<sup>1</sup> Max Barnish,<sup>1</sup> David Packman,<sup>1</sup> Sophie Dodman,<sup>1</sup> Chris Cooper,<sup>1</sup> Tristan Snowsill,<sup>1,2</sup> Tracey Kay,<sup>3</sup> Neil Liversedge,<sup>3</sup> Michelle Parr,<sup>4</sup> Lisa Knight,<sup>3</sup> Chris Hyde,<sup>1</sup> Andrew Shennan<sup>5,6</sup> and Martin Hoyle<sup>1</sup> on behalf of the Peninsula Technology Assessment Group (PenTAG)

<sup>1</sup>Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, University of Exeter, Exeter, UK

<sup>2</sup>Health Economics Group, University of Exeter Medical School, University of Exeter, Exeter, UK

<sup>3</sup>Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

<sup>4</sup>Central Manchester University Hospital NHS Foundation Trust, Manchester, UK

<sup>5</sup>Department of Women and Children's Health, King's College London, London, UK

<sup>6</sup>Guy's and St Thomas' Hospital, London, UK

\*Corresponding author [j.varley-campbell@exeter.ac.uk](mailto:j.varley-campbell@exeter.ac.uk)

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

**Objectives:** The purpose of this report is to assess the test accuracy, clinical effectiveness and cost-effectiveness of the diagnostic tests PartoSure™ (Parsagen Diagnostics Inc., Boston, MA, USA), Actim® Partus (Medix Biochemica, Espoo, Finland) and the Rapid Fetal Fibronectin (fFN)® 10Q Cassette Kit (Hologic, Inc., Marlborough, MA, USA) at thresholds  $\neq$  50 ng/ml [quantitative fFN (qfFN)] for women presenting with signs and symptoms of preterm labour relative to fFN at 50 ng/ml.

**Methods:** Systematic reviews of the published literature were conducted for diagnostic test accuracy (DTA) 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 structure of the model that informed the 2015 National Institute for Health and Care Excellence guidelines on preterm labour diagnosis and treatment, but with antenatal steroids use, as opposed to tocolysis, driving health outcomes.

**Results:** Twenty studies were identified evaluating DTA against the reference standard of delivery within 7 days and seven studies were identified evaluating DTA against the reference standard of delivery within 48 hours. 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, whereas the other indicated little difference between PartoSure and Actim Partus. No study assessing qfFN and PartoSure in the same population was identified. The test accuracy results from the other included studies revealed a high level of uncertainty, primarily attributable to substantial methodological, clinical and statistical heterogeneity between studies. No study compared all three tests simultaneously. 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 with a 'treat-all' strategy, relative to testing with qualitative fFN. Because of the lack of data, our de novo model involved the assumption that management of pregnant women fully adhered to the results of the tests. In the base-case analysis for a woman at 30 weeks' gestation, Actim Partus had lower health-care costs and fewer quality-adjusted life-years (QALYs) than qfFN at 50 ng/ml, reducing costs at a rate of £56,030 per QALY lost compared with qfFN at 50 ng/ml. PartoSure is less costly than Actim Partus while being equally effective, but this is based on diagnostic accuracy data from a small study. Treatment with qfFN at 200 ng/ml and 500 ng/ml resulted in lower cost savings per QALY lost relative to fFN at 50 ng/ml than treatment with Actim Partus. In contrast, qfFN at 10 ng/ml increased QALYs, by 0.002, and had a cost per QALY gained of £140,267 relative to fFN at 50 ng/ml. Similar qualitative results were obtained for women presenting at different gestational ages.

**Conclusion:** There is a high degree of uncertainty surrounding the test accuracy and cost-effectiveness results. We are aware of four ongoing UK trials, two of which plan to enrol > 1000 participants. The results of these trials may significantly alter the findings presented here.

**Study registration:** The study is registered as PROSPERO CRD42017072696.

**Funding:** The National Institute for Health Research Health Technology Assessment programme.

# Contents

<b>List of tables</b>	<b>xiii</b>
<b>List of figures</b>	<b>xvii</b>
<b>Glossary</b>	<b>xix</b>
<b>List of abbreviations</b>	<b>xxiii</b>
<b>Plain English summary</b>	<b>xxv</b>
<b>Scientific summary</b>	<b>xxvii</b>
<b>Chapter 1 Background and definition of the decision problem(s)</b>	<b>1</b>
Conditions and aetiologies	1
<i>Aetiology, pathology and prognosis</i>	1
<i>Epidemiology</i>	1
<i>Incidence and/or prevalence</i>	2
<i>Impact of the health problem</i>	2
Current guidelines	2
Description of the technologies under assessment	4
<i>Summary of the technologies</i>	4
<i>Population</i>	4
<i>Current usage in the NHS</i>	4
<i>Anticipated costs associated with the intervention</i>	7
Comparators	7
<i>Fetal fibronectin used with a threshold of 50 ng/ml</i>	7
<i>Clinical assessment of symptoms alone</i>	8
Care pathways	8
<i>Tocolytic therapy</i>	8
<i>Antenatal corticosteroids</i>	8
<i>Magnesium sulphate</i>	9
Outcomes	9
<b>Chapter 2 Assessment of test accuracy</b>	<b>11</b>
Methods for reviewing test accuracy	11
<i>Methods of the systematic review</i>	11
Results of the systematic review	14
<i>Overview of the quantity and quality of research available</i>	14
<i>Description of the included studies</i>	15
<i>Description of included participants</i>	26
<i>Summary of any treatments given</i>	27
<i>Quality appraisal of included studies</i>	29
<i>Results of quantitative data synthesis (test accuracy data)</i>	34
Summary	42
<i>Data derived from the systematic review of diagnostic test accuracy</i>	42
<i>Summary of the data available across the systematic review and overview</i>	44
<i>Summary of test accuracy data across the systematic review</i>	45

<b>Chapter 3 Assessment of clinical effectiveness (end-to-end) studies</b>	<b>47</b>
Methods for reviewing effectiveness	47
<i>Identification of studies</i>	47
<i>Inclusion and exclusion criteria</i>	47
<i>Other methods</i>	48
Results	48
<i>Quantity and quality of research available</i>	48
<i>Assessment of effectiveness</i>	48
Summary	49
<b>Chapter 4 Data informing the economic modelling</b>	<b>51</b>
<b>Chapter 5 Systematic review of existing cost-effectiveness evidence</b>	<b>53</b>
Methods for reviewing economic evaluation studies	53
<i>Identification of studies</i>	53
<i>Inclusion and exclusion criteria</i>	53
<i>Data extraction strategy</i>	53
<i>Critical appraisal strategy</i>	53
<i>Methods of data synthesis</i>	53
Results	53
Economic evaluation studies	54
<i>Critique</i>	56
<i>Summary</i>	58
Observational cost-minimisation studies	58
Model-based studies	58
<i>The National Institute for Health and Care Excellence 2015 guidelines model</i>	61
Discussion and further research	67
<b>Chapter 6 Independent economic assessment</b>	<b>69</b>
Methods	69
<i>Model structure</i>	69
<i>Population</i>	71
<i>Interventions and comparators</i>	72
<i>Perspective, time horizon and discounting</i>	72
<i>Model parameters</i>	72
<i>Steroid treatment</i>	78
<i>Base-case analyses</i>	85
<i>Scenario analyses</i>	85
<i>Probabilistic sensitivity analyses</i>	85
Results	85
<i>Base-case results</i>	85
<i>Tornado analysis</i>	87
<i>Probabilistic sensitivity analyses</i>	90
<i>Scenario analyses</i>	91
<b>Chapter 7 Assessment of factors relevant to the NHS and other parties</b>	<b>93</b>
<b>Chapter 8 Patient and public involvement</b>	<b>95</b>
<b>Chapter 9 Discussion</b>	<b>97</b>
Review of test accuracy evidence	97
<i>Strengths</i>	98
<i>Weaknesses</i>	98
<i>Areas of uncertainty</i>	99

Review of clinical effectiveness evidence	100
<i>Strengths</i>	100
<i>Weaknesses</i>	100
<i>Areas of uncertainty</i>	101
Review of cost-effectiveness evidence	101
<i>Strengths</i>	101
<i>Weaknesses</i>	101
<i>Areas of uncertainty</i>	102
Independent economic assessment	103
<i>Strengths</i>	104
<i>Weaknesses</i>	104
<i>Areas of uncertainty</i>	105
Other relevant factors	106
<b>Chapter 10 Conclusions</b>	<b>107</b>
Suggested research priorities	107
<b>Acknowledgements</b>	<b>109</b>
<b>References</b>	<b>111</b>
<b>Appendix 1 Literature search strategies</b>	<b>123</b>
<b>Appendix 2 Additional details for quality appraisal for the diagnostic test accuracy review</b>	<b>131</b>
<b>Appendix 3 Supplementary citations and data</b>	<b>135</b>
<b>Appendix 4 Additional information on existing cost-effectiveness studies</b>	<b>169</b>
<b>Appendix 5 Supplementary discussion and tables for the systematic review and selection of utilities</b>	<b>177</b>
<b>Appendix 6 Additional cost-effectiveness results</b>	<b>201</b>
<b>Appendix 7 Additional diagnostic test accuracy data on cervical length</b>	<b>215</b>



# List of tables

<b>TABLE 1</b> Summary of index tests	<b>5</b>
<b>TABLE 2</b> Ongoing trials	<b>16</b>
<b>TABLE 3</b> Study characteristics	<b>17</b>
<b>TABLE 4</b> Participant characteristics	<b>21</b>
<b>TABLE 5</b> Quality appraisal: QUADAS-2	<b>30</b>
<b>TABLE 6</b> Calculated diagnostic accuracy parameters against the 7-day reference standard	<b>35</b>
<b>TABLE 7</b> Calculated diagnostic accuracy parameters against the 48-hour delivery reference standard	<b>41</b>
<b>TABLE 8</b> Summary of evidence and relative accuracy against the 7-day reference standard	<b>44</b>
<b>TABLE 9</b> Summary of test accuracy data from the systematic review	<b>45</b>
<b>TABLE 10</b> The included cost-effectiveness study	<b>55</b>
<b>TABLE 11</b> Model parameter values from the NICE guideline model	<b>57</b>
<b>TABLE 12</b> Modelling studies identified in the systematic review of cost-effectiveness evidence	<b>59</b>
<b>TABLE 13</b> Key assumptions in the NICE 2015 guidelines model	<b>64</b>
<b>TABLE 14</b> The CHEERS checklist for the NICE guidelines model	<b>65</b>
<b>TABLE 15</b> Diagnostic test options compared in cost-effectiveness analyses and their sources	<b>73</b>
<b>TABLE 16</b> Diagnostic accuracy values used in the Peninsula Technology Assessment Group model	<b>74</b>
<b>TABLE 17</b> Baseline risk of neonatal adverse events	<b>77</b>
<b>TABLE 18</b> Treatment effect of ACSs	<b>79</b>
<b>TABLE 19</b> Estimated effectiveness values of steroid treatment used in the model	<b>79</b>
<b>TABLE 20</b> Utilities selected for the economic model	<b>81</b>
<b>TABLE 21</b> Cost parameter values (2016 prices)	<b>83</b>
<b>TABLE 22</b> The PenTAG model specifications and assumptions	<b>86</b>

<b>TABLE 23</b> Summary of ICERs for women presenting at 30 weeks' gestation (at a level 2 hospital)	<b>87</b>
<b>TABLE 24</b> Fully incremental analysis of ICERs for women presenting at 30 weeks' gestation at a level 2 hospital	<b>88</b>
<b>TABLE 25</b> Breakdown of base-case results (discounted costs and QALYs)	<b>89</b>
<b>TABLE 26</b> Incremental cost-effectiveness ratios vs. fFN at 50 ng/ml	<b>92</b>
<b>TABLE 27</b> Included citations	<b>136</b>
<b>TABLE 28</b> Alere Inc. (Actim Partus) submitted citations	<b>138</b>
<b>TABLE 29</b> Hologic, Inc. (fFN) submitted citations	<b>140</b>
<b>TABLE 30</b> Parsagen Diagnostics Inc. (PartoSure) submitted citations	<b>143</b>
<b>TABLE 31</b> Studies excluded at full-text review, with reasons	<b>145</b>
<b>TABLE 32</b> The DTA values for prediction of delivery within 48 hours	<b>166</b>
<b>TABLE 33</b> The DTA values for prediction of delivery within 7 days	<b>167</b>
<b>TABLE 34</b> Parameter values in Chuck and Nguyen	<b>171</b>
<b>TABLE 35</b> Model parameter values in Deshpande <i>et al.</i>	<b>172</b>
<b>TABLE 36</b> Key parameters from Boyd <i>et al.</i>	<b>172</b>
<b>TABLE 37</b> Effectiveness parameters based on Mozurkewich <i>et al.</i>	<b>174</b>
<b>TABLE 38</b> Effectiveness parameters based on van Baaren <i>et al.</i>	<b>175</b>
<b>TABLE 39</b> Summary of papers from the systematic search, with information on quality of life for preterm children	<b>178</b>
<b>TABLE 40</b> Secondary sources for utilities (referenced by papers found in the systematic search)	<b>182</b>
<b>TABLE 41</b> Summary of papers on quality of life for mothers	<b>184</b>
<b>TABLE 42</b> Summary of utilities for preterm children	<b>185</b>
<b>TABLE 43</b> Summary of utilities from secondary sources cited by studies found in the systematic search	<b>189</b>
<b>TABLE 44</b> Mapped utilities for the quality of life of mothers	<b>192</b>
<b>TABLE 45</b> Random-effects GLS estimates of utility on birthweight	<b>196</b>
<b>TABLE 46</b> Total discounted QALYs for preterm survivors, by gestational age	<b>198</b>



<b>TABLE 47</b> Total discounted QALY loss for RDS and IVH as used in the economic model	199
<b>TABLE 48</b> Total discounted QALYs for mothers (assuming age at birth of 30 years)	199
<b>TABLE 49</b> Summary of ICERs for women presenting at 26 weeks' gestation (level 2 hospital)	201
<b>TABLE 50</b> Fully incremental analysis of ICERs for women presenting at 26 weeks' gestation (level 2 hospital)	202
<b>TABLE 51</b> Breakdown of discounted costs and QALYs for women presenting at 26 weeks' gestation	202
<b>TABLE 52</b> Summary of ICERs for women presenting at 33 weeks' gestation (level 2 hospital)	203
<b>TABLE 53</b> Fully incremental analysis of ICERs for women presenting at 33 weeks' gestation (level 2 hospital)	204
<b>TABLE 54</b> Breakdown of discounted costs and QALYs for women presenting at 33 weeks' gestation	204
<b>TABLE 55</b> Results for Actim Partus and no testing vs. fFN at 50 ng/ml using data from Cooper <i>et al.</i> for women presenting at 30 weeks' gestation (level 2 hospital)	209
<b>TABLE 56</b> Results for fFN (various thresholds) vs. fFN at 50 ng/ml using data from Abbott <i>et al.</i> for women presenting at 30 weeks' gestation (level 2 hospital)	209
<b>TABLE 57</b> Results for no testing and Actim Partus vs. fFN at 50 ng/ml using data from meta-analysis for women presenting at 30 weeks' gestation (level 2 hospital)	211
<b>TABLE 58</b> Incremental cost-effectiveness ratios vs. fFN at 50 ng/ml for women presenting at 26 weeks' gestation (level 2 hospital)	212
<b>TABLE 59</b> Incremental cost-effectiveness ratios vs. fFN at 50 ng/ml for women presenting at 33 weeks' gestation (level 2 hospital)	213
<b>TABLE 60</b> 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	214
<b>TABLE 61</b> 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	214
<b>TABLE 62</b> Test accuracy results (against preterm birth within 7 days) for index tests, cervical length measurement and fFN at 50 ng/ml	216
<b>TABLE 63</b> Test accuracy results (against preterm birth within 7 days) for index tests and fFN at 50 ng/ml	218

<b>TABLE 64</b> Summary of evidence and relative accuracy against the 7-day reference standard	<b>223</b>
<b>TABLE 65</b> Summary table	<b>225</b>

# List of figures

<b>FIGURE 1</b> Diagnosis of preterm labour from section 9 of the 2015 NICE guidance on preterm labour and birth	<b>3</b>
<b>FIGURE 2</b> Summary of the selection process	<b>15</b>
<b>FIGURE 3</b> The ROC plot for Actim Partus against the 7-day reference standard	<b>38</b>
<b>FIGURE 4</b> The ROC plot for PartoSure against the 7-day reference standard	<b>39</b>
<b>FIGURE 5</b> The ROC plot for Actim Partus against the 48-hour delivery reference standard	<b>42</b>
<b>FIGURE 6</b> The PRISMA flow diagram for the economic evaluation review	<b>54</b>
<b>FIGURE 7</b> The NICE 2015 guidelines model	<b>62</b>
<b>FIGURE 8</b> Baseline (without tocolytics) risks in the NICE guideline model	<b>63</b>
<b>FIGURE 9</b> The Peninsula Technology Assessment Group cost-effectiveness model structure of diagnosis and treatment of women with symptoms of preterm labour	<b>70</b>
<b>FIGURE 10</b> Empirical summary ROC points across evaluated studies (7-day)	<b>76</b>
<b>FIGURE 11</b> Baseline risks of neonatal adverse events	<b>77</b>
<b>FIGURE 12</b> Total lifetime QALYs for preterm children by gestational age and health state, using age and survival adjustment (solid lines) and NICE's unadjusted method (dashed lines)	<b>82</b>
<b>FIGURE 13</b> Incremental costs and benefits of index tests against comparator (fFN at 50 ng/ml)	<b>88</b>
<b>FIGURE 14</b> Tornado diagram for model parameters with the most influential variation in value: PartoSure	<b>90</b>
<b>FIGURE 15</b> Probabilistic analysis: women presenting at 30 weeks' gestation	<b>90</b>
<b>FIGURE 16</b> Decision tree of fFN testing strategy in Chuck and Nguyen	<b>170</b>
<b>FIGURE 17</b> Model of Rapid fFN testing from Mozurkewich <i>et al.</i>	<b>173</b>
<b>FIGURE 18</b> Model of fFN diagnostic testing in van Baaren <i>et al.</i>	<b>175</b>
<b>FIGURE 19</b> Tornado diagram of most influential model parameter values: treat all	<b>205</b>
<b>FIGURE 20</b> Tornado diagram of most influential model parameter values: fFN at 10 ng/ml	<b>206</b>

<b>FIGURE 21</b> Tornado diagram of most influential model parameter values: fFN at 200 ng/ml	<b>206</b>
<b>FIGURE 22</b> Tornado diagram of most influential model parameter values: fFN at 500 ng/ml	<b>207</b>
<b>FIGURE 23</b> Tornado diagram of most influential model parameter values: Actim Partus	<b>207</b>
<b>FIGURE 24</b> Probabilistic analysis: women presenting at 33 weeks' gestation	<b>208</b>
<b>FIGURE 25</b> Probabilistic analysis: women presenting at 26 weeks' gestation	<b>208</b>

# Glossary

**Antenatal corticosteroid therapy** Therapy administered to women when preterm delivery is anticipated, to enhance fetal lung maturation. The aim of treatment is to prevent respiratory distress syndrome and reduce mortality and morbidity among preterm infants.

**Bronchopulmonary dysplasia** A chronic lung disease that affects premature infants requiring oxygen therapy. It commonly occurs secondary to respiratory distress syndrome.

**Cervical length** Cervical length measurement via transvaginal ultrasound scan 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 a higher risk of preterm delivery.

**Cervical os** Opening of the uterine cervix (anatomy). It 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** 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 cervicovaginal secretions earlier than 35 weeks can be used to predict onset of preterm delivery (fetal fibronectin 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 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 through the administration of labour-inducing drugs or delivery via caesarean section.

**Incremental cost-effectiveness ratio** A term used in health economics to compare the difference in the cost and the effectiveness of two interventions/tests:

$$\text{Incremental cost-effectiveness ratio} = (C_1 - C_0)/(E_1 - E_0), \quad (\text{a})$$

where  $C_1$  = cost of intervention,  $C_0$  = cost of control,  $E_1$  = effectiveness of intervention and  $E_0$  = effectiveness of control.

**Intraventricular haemorrhage** A condition associated with preterm delivery, characterised by bleeding into the ventricles of the brain. Severity is categorised by four grades: grades 1 and 2 denote a smaller amount of bleeding and grades 3 and 4 denote more severe bleeding.

**Likelihood ratio** In this study, the likelihood of a given test result in a patient who has a preterm delivery compared with the likelihood of the same result in a patient who does not deliver preterm.

Positive likelihood ratio: how much more often a positive test result occurs in people who deliver preterm than in those who do not:

$$\text{Positive likelihood ratio} = \frac{P(\text{Test + ve} \mid \text{preterm})}{P(\text{Test + ve} \mid \text{not preterm})} = \frac{\text{Sensitivity}}{1 - \text{Specificity}}. \quad (\text{b})$$

Negative likelihood ratio: how much less likely a negative result is in people with preterm delivery than in those without preterm delivery:

$$\text{Negative likelihood ratio} = \frac{P(\text{Test - ve} \mid \text{preterm})}{P(\text{Test - ve} \mid \text{not preterm})} = \frac{1 - \text{Sensitivity}}{\text{Specificity}}. \quad (\text{c})$$

**Meta-analysis** A statistical technique that combines data from various studies evaluating the same index test to calculate pooled diagnostic accuracy estimates.

**Multiple gestation pregnancy** A pregnancy in which the number of fetuses exceeds one.

**Negative predictive value** In this study, the proportion of people with a negative result who will not deliver preterm (within 48 hours or 7 days):

$$\text{Negative predictive value} = \text{true negative} / (\text{true negative} + \text{false negative}). \quad (\text{d})$$

**Parity** The number of times a woman has carried a pregnancy to a viable gestation.

**Phosphorylated insulin-like growth factor-binding protein-1** A 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 (Medix Biochemica, Espoo, Finland)].

**Placental alpha microglobulin-1** This 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™ (Parsagen Diagnostics Inc., Boston, MA, USA) test].

**Positive predictive value** In this study, the proportion of people with a positive result who will deliver preterm (within 48 hours or 7 days):

$$\text{Positive predictive value} = \text{true positive} / (\text{true positive} + \text{false positive}). \quad (\text{e})$$

**Preterm birth/delivery** The delivery of a live baby before 37<sup>+0</sup> weeks' gestational age: < 28 weeks' gestational age = extremely preterm, ≥ 28 weeks' and < 32 weeks' gestational age = very preterm and ≥ 32 weeks' and < 37 weeks' gestational age = moderate to late preterm.

**Preterm premature rupture of membranes** Premature (< 37 weeks' gestation) rupture of the amniotic sac surrounding the fetus before the onset of established labour. Women experiencing preterm premature rupture of membranes are at increased risk of amniotic infection and preterm delivery.

**Prevalence** In this study, the proportion of women actually delivering preterm (within 48 hours or 7 days).

**Quality-adjusted life-year** A measure of disease burden that combines length and quality of life.

**Receiver operating characteristic plot** A graphical depiction of diagnostic test accuracy data for all included studies.

**Reference standard** The best diagnostic test currently available, against which an index test is assessed. Owing to the predictive nature of the index tests in this study, the reference standard for all included studies was preterm delivery within 48 hours or 7 days.

**Respiratory distress syndrome** A breathing disorder that commonly affects premature babies and is attributable to insufficient surfactant production in immature lungs.

**Sensitivity** The ability of a diagnostic test to correctly identify women in whom delivery is imminent (within 48 hours or 7 days):

$$\text{Sensitivity} = \text{true positive} / (\text{true positive} + \text{false negative}). \quad (\text{f})$$

**Single-gate study** A study design in which participants' disease statuses are unknown and the index test result is evaluated against the reference standard to confirm the diagnosis.

**Specificity** In this study, the ability of a diagnostic test to correctly identify women for whom delivery is not imminent (within 48 hours or 7 days).

$$\text{Specificity} = \text{true negative} / (\text{true negative} + \text{false positive}). \quad (\text{g})$$

**Test failure** The rate of non-informative test results.

**Time to test** The 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 time to complete corticosteroid administration; however, this is no longer recommended practice.





## List of abbreviations

ACS	antenatal corticosteroid	IQR	interquartile range
AG	assessment group	ISRCTN	International Standard Randomised Controlled Trial Number
APOSTEL-1	Assessment of Perinatal Outcome after Sustained Tocolysis in Early Labour	IVH	intraventricular haemorrhage
BAPM	British Association of Perinatal Medicine	LR	likelihood ratio
BMI	body mass index	LR–	likelihood ratio for a negative test result
BNF	<i>British National Formulary</i>	LR+	likelihood ratio for a positive test result
CDC	Centers for Disease Control and Prevention	MAPP-QOL	Maternal Postpartum Quality of Life
CENTRAL	Cochrane Central Register of Controlled Trials	MLE	maximum likelihood estimation
CHEERS	Consolidated Health Economic Evaluation Reporting Standards	NHS EED	NHS Economic Evaluation Database
CI	confidence interval	NICE	National Institute for Health and Care Excellence
CINAHL	Cumulative Index to Nursing and Allied Health Literature	NICU	neonatal intensive care unit
DARE	Database of Abstracts of Reviews of Effects	NNAP	National Neonatal Audit Programme
DTA	diagnostic test accuracy	NPV	negative predictive value
ELISA	enzyme-linked immunosorbent assay	OLS	ordinary least squares
EPICE	Effective Perinatal Intensive Care in Europe	ONS	Office for National Statistics
EQ-5D	EuroQol-5 Dimensions	OR	odds ratio
fFN	fetal fibronectin	PAMG-1	placental alpha microglobulin-1
FPR	false-positive rate	PenTAG	Peninsula Technology Assessment Group
GBP	Great British pounds	ph(IGFBP-1)	phosphorylated insulin-like growth factor-binding protein-1
GLS	generalised least squares	PPROM	Preterm Premature Rupture of Membranes
HRG	Healthcare Resource Group	PPV	positive predictive value
HTA	Health Technology Assessment	PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
HUI	Health Utilities Index	PVL	periventricular leukomalacia
ICER	incremental cost-effectiveness ratio	QALY	quality-adjusted life-year
ICH	intracranial haemorrhage	qfFN	quantitative fetal fibronectin
ICU	intensive care unit	RCOG	Royal College of Obstetricians and Gynaecologists

RCT	randomised controlled trial	SF-36	Short Form questionnaire-36 items
RDS	respiratory distress syndrome	VAT	value-added tax
ROC	receiver operating characteristic	WHO	World Health Organization
SD	standard deviation		

### Note

This monograph is based on the Technology Assessment Report produced for NICE. The full report contained a considerable number of data that were deemed confidential. The full report was used by the Appraisal Committee at NICE in their deliberations. The full report with each piece of confidential data removed and replaced by the statement 'confidential information (or data) removed' is available on the NICE website: [www.nice.org.uk](http://www.nice.org.uk).

The present monograph presents as full a version of the report as is possible while retaining readability, but some sections, sentences, tables and figures have been removed. Readers should bear in mind that the discussion, conclusions and implications for practice and research are based on all the data considered in the original full NICE report.

## Plain English summary

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 could help the doctor to decide whether the mother needs treatment or to move to a specialist hospital or if she could be sent home (if it is a false alarm).

Our report compares three tests [PartoSure™ (Parsagen Diagnostics Inc., Boston, MA, USA), Actim® Partus (Medix Biochemica, Espoo, Finland) and the Fetal Fibronectin (fFN) Test (Hologic, Inc., Marlborough, MA, USA)] on how well they predict an early birth and how the costs and the long-term health outcomes of the child compare between and among tests.

All the published literature reporting the accuracy of the three tests and their costs was reviewed.

We developed a new cost-effectiveness model, which estimated the long-term health outcomes of the child based on the test results.

Twenty of the studies reviewed looked at how good the tests were at predicting an early birth within the next 7 days, and six looked at predicting birth within 48 hours. 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 fFN compared with Actim Partus. PartoSure appears 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. There were no data that allowed us to compare all three tests together.

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



# Scientific summary

## Background

Preterm (premature) birth, as defined by the World Health Organization, refers to birth of alive baby before 37 weeks of gestation. Approximately 8% of births in England and Wales are premature. Preterm birth can result in serious short-term health issues for the infant, including difficulties with breathing [respiratory distress syndrome (RDS)] and feeding and 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 National Institute for Health and Care Excellence (NICE) guidelines (published in 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  $\leq 29^{+6}$  weeks pregnant, treatment for preterm labour is recommended. If the clinical assessment suggests that the woman is in suspected preterm labour and she is  $\geq 30^{+0}$  weeks pregnant, then the following tests should be conducted:

1. a transvaginal ultrasound scan measurement of cervical length (positive if  $< 15$  mm)
2. if transvaginal ultrasound scan measurement of cervical length is unavailable or unacceptable, a fetal fibronectin (fFN) test (positive if concentration is  $\geq 50$  ng/ml).

Accurate diagnoses of preterm births could prevent unnecessary (or ensure appropriate) admissions into hospitals 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 effectiveness and cost-effectiveness:

1. PartoSure™ (Parsagen Diagnostics Inc., Boston, MA, USA) – a point-of-care dipstick test that detects placental alpha microglobulin-1 (PAMG-1) in vaginal secretions
2. Actim® Partus (Medix Biochemica, Espoo, Finland; distributed by Alere Inc.) – a point-of-care dipstick test that detects phosphorylated insulin-like growth factor-binding protein-1 [ph(IGFBP-1)] in cervical secretions
3. Rapid fFN® 10Q Cassette Kit (Hologic, Inc., Marlborough, MA, USA) [referred to in this report as quantitative fFN (qfFN)] used with a threshold  $\neq 50$  ng/ml – a point-of-care quantitative test that detects the concentration of fFN in cervicovaginal fluid.

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

- diagnostic test accuracy (DTA) 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 was conducted.

## Review of test accuracy

### Methods

A systematic review was undertaken to assess the DTA of PartoSure, Actim Partus and qfFN. Studies were identified by searching seven bibliographic databases, searching 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 and qfFN at thresholds  $\neq$ 50 ng/ml
- reference standards – preterm delivery within 48 hours or within 7 days
- comparators – clinical assessment of symptoms alone, qualitative fFN, or qfFN at a threshold of 50 ng/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 the inclusion criteria at the 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. When the data allowed, summary receiver operating characteristic plots were generated and meta-analyses were conducted.

### Results

Twenty studies met the inclusion criteria: 16 studies assessed Actim Partus, four assessed PartoSure and two assessed qfFN.

Sufficient evidence for pooling the test accuracy data was available only for Actim Partus and PartoSure against the 7-day reference standard and for Actim Partus against the 48-hour 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 tests were those that assessed two or more different tests within the same population. We identified two such studies. In the first study, depending on the threshold used, qfFN was more or less sensitive and specific than Actim Partus. In the second study, there was little difference between the sensitivity and specificity of PartoSure and Actim Partus. No studies assessed qfFN 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 had sensitivity of 94.7% [95% confidence interval (CI) 89.9% to 97.7%] and specificity of 92.4% (95% CI 88.9% to 95.1%), whereas the study reporting the worst results had sensitivity of 33.3% (95% CI 4.3% to 77.7%) and specificity of 74.1% (95% CI 69.1% to 78.6%). For PartoSure ( $n = 4$  studies), the study with the best overall sensitivity and specificity results had sensitivity of 100.0% (95% CI 73.5% to 100.0%) and specificity of 95.4% (95% CI 88.6% to 98.7%). The study reporting the worst results had sensitivity of 0.0% (95% CI 0.0% to 97.5%) and specificity of 97.5% (95% CI 96.8% to 99.9%); the low sensitivity reported in that study is attributable to only one woman delivering preterm (within 7 days) and her testing (falsely) negative within the study sample of size 41. fFN at a threshold of 10 ng/ml ( $n = 2$  studies) had a sensitivity range of 93.8% (95% CI 82.8% to 98.7%) to 95.7% (95% CI 87.8% to 99.1%) and a specificity range of 32.2% (95% CI 27.7% to 37.0%) to 42.3% (95% CI 36.5% to 48.4%), at a threshold of 200 ng/ml, sensitivity ranged from 70.8% (95% CI 55.9% to 83.0%) to 71.0% (95% CI 58.8% to 81.3%) and specificity ranged from 78.6% (95% CI 74.3% to 82.5%) to 83.6% (95% CI 78.8% to 87.8%), and at a threshold of 500 ng/ml, sensitivity ranged from 29.2% (95% CI 17.0% to 44.1%) to 42.0% (95% CI 30.2% to 54.5%) and specificity ranged from 94.3% (95% CI 91.6% to

96.4%) to 95.7% (95% CI 92.7% to 97.8%). Given the large ranges between studies assessing the same test and the significant overlapping of CIs, 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 ( $n = 6$  studies) and PartoSure ( $n = 1$  study) against the 48-hour reference standard, because no studies were identified that assessed qfFN. 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 be premature to attempt to deduce which test was superior against the 48-hour 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 DTA to identify randomised controlled or controlled studies of the tests (PartoSure, Actim Partus or fFN at thresholds  $\neq 50$  ng/ml). Studies could compare the tests with each other or with fFN at a threshold of 50 ng/ml, or with 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 qfFN. 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; that study modelled the cost-effectiveness of a 'treat-all' strategy, relative to testing with qualitative fFN to determine treatment. Based on the findings of that study, we calculated that the incremental cost-effectiveness ratio (ICER) of treating all suspected cases of preterm labour with antenatal corticosteroids (ACSS) is £20,942 per quality-adjusted life-year (QALY) gained.

This identified study also compared the use of four different qfFN thresholds (10, 50, 200 and 500 ng/ml). Based on the results, we also calculated that testing at 200 ng/ml dominates testing at lower thresholds, owing to treatment and health-care costs saved. However, the ICER of testing at 200 ng/ml, relative to a higher threshold of 500 ng/ml, was found to be £10,415 per QALY gained. Therefore, our calculations may support the study authors' conclusions that using a 200-ng/ml threshold for qfFN was the optimal testing threshold. However, owing 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 reviewed. 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 for 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 quantity or quality of life). The first was a Canadian study that found that testing with fFN added total costs of approximately US\$4M, relative to no testing. The second was a UK study that compared clinical examination alone with clinical examination with a fFN test. This study found that using fFN saved the NHS £23.88 per patient, with the additional test costs offset by the savings in hospital resource costs being resulting with from treating fewer women.

Three studies provided cost-effectiveness analyses. The first (in the UK) compared testing with fFN with a 'treat-all' strategy. This model was unique in allowing for < 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 (in the USA) found that treating all patients had incremental costs of US\$433,000 per case of RDS avoided and US\$1,300,000 per neonatal life saved relative to fFN (1999 prices). It differed from other models by explicitly modelling preterm birth within 48 hours of testing. The third study (in the 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 conclude that it provides the most suitable structure for the decision problem on which to base our own model.

## Independent economic assessment

We developed a new model that adopted the best published methodological practice including that of the 2015 NICE guidelines model. It models diagnostic outcomes as a decision tree structure and projects long-term health outcomes many years into the future based on the occurrence of major neonatal adverse events. Unlike the NICE model, which assumed that all treatment involved tocolysis, our model considers treatment with ACSs only. Use of tocolysis is only assumed in case of hospital transfer. This is based on both 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 hospital with a lower neonatal unit to a higher-level unit hospital in mothers of the youngest gestational ages
- accounting for the costs and benefits of ACSs for treatment of preterm labour, and the cost of tocolysis for transfers
- using gestational age of birth-specific inpatient costs estimated from national registry data on level of care received by newborn premature infants until hospital discharge.

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 scan is not available or acceptable, is tested using one of fFN, Actim Partus or PartoSure. Regardless of the result, this woman can:

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



If a woman tests positive, she is treated with steroids. If the gestational age is < 28 weeks, and she presents at a hospital with a level 1 or 2 neonatal unit, 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. Owing to a lack of suitable data in the literature, we used proxy utility values for IVH and RDS. Because only one study provided data for the quality of life of mothers who had 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 model that informed the existing 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.

As there was no study that compared all the diagnostic options, we produced an economic assessment for the individual comparative studies separately. The results from our base-case analysis (for a woman presenting at 30 weeks' gestation) are as follows. Using test accuracy data from one study, 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 a £56,030 cost saving per QALY lost versus 50 ng/ml of fFN. Using test accuracy data from another study, we find that PartoSure is less costly than Actim Partus while being equally effective. Indirectly comparing PartoSure with 50 ng/ml fFN (using data from two studies) yields a saving of £81,922 per QALY loss with the former relative to the latter test. This estimate is highly uncertain given the indirect comparison source and the small size of one of the studies used. Furthermore, qfFN at the 200 ng/ml and 500 ng/ml thresholds saves £25,209 and £17,025 per QALY loss, respectively, relative to 50 ng/ml of fFN. qfFN at 10 ng/ml was the only test option that increased QALYs, by 0.002, relative to 50 ng/ml fFN, and had an incremental cost per QALY gained of £140,267. The discounted QALY differences between new test options against 50 ng/ml fFN were all smaller than 0.03.

## 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 accuracy studies and the lack of any study of the tests on decision-making and clinical outcomes. 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 > 1000 participants, the results of which are likely to affect these conclusions.

## Study registration

The study is registered as PROSPERO CRD42017072696.

## Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.



# Chapter 1 Background and definition of the decision problem(s)

## Conditions and aetiologies

Preterm (premature) birth, as defined by the World Health Organization (WHO), refers to babies born alive before 37 weeks and 0 days of gestation (37<sup>+0</sup> weeks).<sup>1</sup>

Preterm birth can be serious for an infant in terms of both short- and long-term health problems and an increased risk of mortality. For example, short-term problems include difficulties with breathing [respiratory distress syndrome (RDS)] and feeding and an increased risk of infections and bleeding within the brain [intraventricular haemorrhages (IVHs)]. Meanwhile, long-term problems include an increased risk of cerebral palsy, cognitive and visual impairment and respiratory illnesses.<sup>2,3</sup>

### Aetiology, pathology and prognosis

The WHO<sup>1</sup> subcategorises preterm birth based on gestational age as:

- extremely preterm – < 28 weeks' gestational age
- very preterm – ≥ 28 weeks' and < 32 weeks' gestational age
- moderate to late preterm – ≥ 32 weeks' and < 37 weeks' gestational age.

Iatrogenic preterm births are medically instigated deliveries, such as early labour induction or caesarean section.<sup>4</sup> These elective deliveries aim to reduce health risks to the mother or fetus owing to complications such as hypertension, intrauterine growth restriction or pre-eclampsia.<sup>4</sup>

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).<sup>5</sup> Spontaneous preterm deliveries can be broadly categorised as either spontaneous labour with intact membranes or those following preterm premature rupture of membranes (PPROMs).<sup>4</sup> Factors associated with an increased risk of preterm delivery include stress, tobacco use, drug abuse, trauma, multifetal gestations, in vitro fertilisation, low body mass index (BMI) before pregnancy, extremes of maternal age, diabetes mellitus, high blood pressure and infection.<sup>6,7</sup> However, previous preterm delivery is the greatest risk factor for preterm birth.<sup>8</sup>

Symptoms of suspected preterm labour include painful contractions or cramps, abdominal and low-back pain and an increase or change in vaginal discharge.<sup>9</sup> 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 > 90% of women presenting with symptoms of preterm labour do not go on to deliver in the next 2 weeks and, of these, 50% will continue with pregnancy until full term.<sup>10,11</sup> It is important to determine whether or not preterm labour is the cause of the symptoms and to assess the risk of preterm delivery to allow appropriate management to begin as soon as possible.<sup>12</sup>

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

### Epidemiology

Data from the England and Wales 2016 birth cohort<sup>13</sup> report 54,143 live, preterm deliveries in accordance with the WHO definition of preterm birth (< 37 weeks' gestational age), corresponding to 7.8% of all live births. Of these deliveries, 5.9% were categorised as extremely preterm (< 28 weeks' gestation), 10.4%

were very preterm (gestational age of  $\geq 28$  to  $< 32$  weeks) and 83.7% were moderate to late preterm ( $\geq 32$  to  $< 37$  weeks' gestation).<sup>13</sup>

The 2016 UK birth cohort data collected by the Office for National Statistics (ONS)<sup>13</sup> show that the rate of preterm births varies between ethnic populations, with the highest proportion of preterm births occurring in black Caribbean and Indian populations (10.4% and 8.03% of pregnancies in these populations, respectively) and the lowest rate of preterm births occurring in women of 'white other' ethnicity (6.6%). The rate of preterm delivery in the population in which ethnicity was 'not stated' was 8.3%.<sup>13</sup> In the UK, preterm labour, particularly extreme preterm labour, disproportionately affects women from low socioeconomic backgrounds.<sup>14,15</sup>

### **Incidence and/or prevalence**

Improvements in perinatal health-care services have resulted in vastly improved outcomes for babies born preterm, yet the prevalence of preterm birth continues to rise.<sup>1,16</sup>

Preterm birth rates vary between countries, with higher prevalence and poorer outcomes in lower-income countries.<sup>16</sup> However, preterm birth is a global issue that also affects developed countries.

### **Impact of the 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 aged  $< 5$  years.<sup>16,17</sup>

Morbidities associated with preterm birth are both acute and chronic and can affect all organ systems. Respiratory distress can progress to bronchopulmonary dysplasia<sup>18</sup> and cerebral pathology (e.g. IVHs and ischaemia can lead to neurodevelopmental disorders including learning and behavioural difficulties).<sup>19,20</sup> In addition, gastrointestinal disorders and immunodeficiencies are also associated with preterm birth.<sup>21,22</sup>

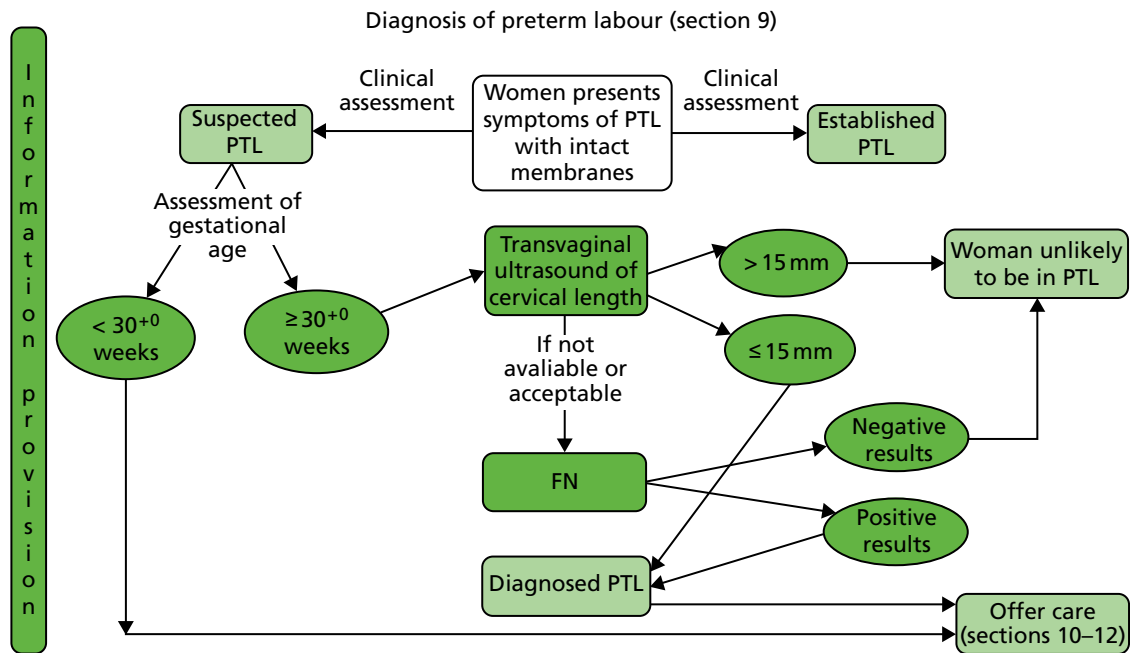
Although mortality and morbidity rates are higher for infants delivered at lower gestational ages and lower birthweights, near-term premature infants remain at a considerably higher risk of complications than their full-term counterparts.<sup>20</sup>

Preterm deliveries place a significant cost burden on the NHS. In addition to initial hospitalisation, rehospitalisation and rehabilitation, other direct medical costs include medication, aids and devices such as wheelchairs, visits to physicians and home care.<sup>23</sup> Direct non-medical costs such as special education, adaptations to homes or cars, special meal requirements, higher insurance premiums and other disease-associated costs are an expensive burden on both families and the state.<sup>23</sup>

## **Current guidelines**

The National Institute for Health and Care Excellence (NICE) guideline<sup>24</sup> on preterm labour (*Figure 1*) and birth states that women reporting symptoms of preterm labour who have intact membranes should have a clinical assessment that includes:

- 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).



**FIGURE 1** Diagnosis of preterm labour from section 9 of the 2015 NICE guidance on preterm labour and birth.<sup>24</sup> FN, fetal fibronectin; PTL, preterm labour. Reproduced from: Royal College of Obstetricians NICE Guideline 25 Preterm Labour and Birth, London, RCOG, November 2015, with the permission of the Royal College of Obstetricians and Gynaecologists.<sup>24</sup>

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.<sup>24</sup>

If the clinical assessment suggests that the woman is in suspected preterm labour and she is  $\geq 30^{+0}$  weeks pregnant then the following tests should be conducted:<sup>24</sup>

- Transvaginal ultrasound scan 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 the community and advised to return if symptoms reappear.
  - If cervical length is  $\leq 15$  mm, the woman is diagnosed as being in preterm labour and should be offered treatment.
- If transvaginal ultrasound scan 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  $\geq 30^{+0}$  weeks pregnant.
  - If the fFN test result is negative (concentration of  $< 50$  ng/ml), the woman is unlikely to be in preterm labour and could be discharged home with routine follow-up in the community and advised to return if symptoms reappear.
  - If the fFN test result is positive (concentration of  $\geq 50$  ng/ml), the woman is diagnosed as being in preterm labour and should be offered treatment.

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

## Description of the technologies under assessment

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

### 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 50 ng/ml) is currently being used. A summary of information relating to the tests is given in *Table 1*.

#### PartoSure

PartoSure™ (Parsagen Diagnostics Inc., Boston, MA, USA) 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 a 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 that the presence of PAMG-1 in vaginal discharge is predictive of imminent delivery.<sup>25</sup>

#### Actim Partus

Actim® Partus (Medix Biochemica, Espoo, Finland; distributed by Alere Inc.) is a CE-marked qualitative immunochromatographic point-of-care test that detects phosphorylated insulin-like growth factor binding protein-1 [ph(IGFBP-1)] in cervical secretions. ph(IGFBP-1) is made by cells lining the uterus and leaks into the cervix when delivery is imminent.<sup>12</sup>

#### Rapid Fetal Fibronectin 10Q Cassette Kit

The Rapid fFN® 10Q Cassette Kit (Hologic, Inc., Marlborough, MA, USA) 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. fFN is a glycoprotein that connects membranes of the uterus and fetal membranes, which begins to degrade after 35 weeks of pregnancy or soon before preterm birth.

### Population

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

### Identification of important subgroups

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:<sup>24</sup>

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

### Current usage in the NHS

Current NICE guidelines are described in *Current guidelines*.

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. In addition, clinicians advise that transvaginal ultrasound scan is rarely used in routine practice owing to a lack of trained staff available, experience or equipment availability.

TABLE 1 Summary of index tests

Category	Actim® Partus (Medix Biochemica, Espoo, Finland) <sup>26</sup>	PartoSure™ (Parsagen Diagnostics Inc., Boston, MA, USA) <sup>27</sup>	Rapid fFN® 10Q Cassette Kit (Hologic, Inc., Marlborough, MA, USA) <sup>28</sup>
Gestational age	From 22 weeks	From 20 <sup>+0</sup> weeks to 36 <sup>+6</sup> weeks	From 22 <sup>+0</sup> weeks to 35 <sup>+6</sup> weeks
Contraindications	Ruptured membranes, vaginal bleeding (moderate or heavy), amniotic fluid	Significant blood on the swab, within 6 hours of vaginal disinfectant solutions or medicines. Inaccurate results may be likely with previous placenta previa or digital examination or in presence of meconium, antifungal creams, suppositories, lubricants, moisturisers, talcum powder or baby oil	Advanced cervical dilatation ( $\geq 3$ cm), ruptured membranes, cervical cerclage, placental abruption, placenta previa (moderate) or vaginal bleeding (heavy). Inaccurate results may be likely with sexual intercourse, digital cervical examination or transvaginal ultrasound scan and bacteria, bilirubin and semen. A negative test result is still valid if in the presence of semen
Instructions	<ol style="list-style-type: none"> <li>1. Take a 10- to 15-second swab from the cervical os during sterile speculum examination before any other investigations</li> <li>2. Swirl the swab vigorously in the specimen extraction solution for 10–15 seconds</li> <li>3. Place the dipstick in the specimen extraction solution until the liquid reaches the result area</li> <li>4. Remove dipstick from sample solution and lay horizontally for 5 minutes</li> </ol> <p>Negative results (one blue line) should be confirmed at 5 minutes: highly unlikely that patient will deliver within the next 2 weeks</p> <p>Positive results (two blue lines) can be read as soon as it becomes visible (if before 5 minutes). Risk of a preterm delivery is elevated</p>	<ol style="list-style-type: none"> <li>1. Open the solvent vial and place in a vertical position by holding by the cap with all the liquid at the bottom</li> <li>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 (5–7 cm deep) for 30 seconds</li> <li>3. Place the swab in the solvent vial and rinse by rotating for 30 seconds</li> <li>4. Insert test strip into the solvent for 5 minutes</li> </ol> <p>Negative results (one line) should be confirmed at 5 minutes</p> <p>Positive results (two lines) can be read as soon as it becomes visible (if before 5 minutes)</p>	<ol style="list-style-type: none"> <li>1. Perform daily analyser quality control</li> <li>2. During speculum examination, collect swab sample from the posterior fornix of the vagina and transfer to the transfer tube</li> <li>3. Mix sample in transport tube prior to removing swab expressing as much liquid as possible from the swab to tube</li> <li>4. Mix the patient sample by removing the fFN 10Q Cassette from the foil pouch, enter necessary information into analyser and insert the cassette into analyser</li> <li>5. When prompted pipette 200 <math>\mu</math>l of patient sample into sample application</li> <li>6. Wait 10 minutes (7 minutes of incubation and 2–3 minutes of analysis)</li> <li>7. fFN concentration will be displayed</li> </ol>

continued

**TABLE 1** Summary of index tests (*continued*)

Category	Actim® Partus (Medix Biochemica, Espoo, Finland) <sup>26</sup>	PartoSure™ (Parsagen Diagnostics Inc., Boston, MA, USA) <sup>27</sup>	Rapid fFN® 10Q Cassette Kit (Hologic, Inc., Marlborough, MA, USA) <sup>28</sup>
Kit components	<ul style="list-style-type: none"> <li>One sterile polyester swab for specimen collection</li> <li>One tube of specimen extraction solution (0.5 ml). This phosphate-buffered solution contains bovine serum albumin, protease inhibitors and preservatives</li> <li>One dipstick in a sealed aluminium foil pouch with desiccant</li> </ul>	<ul style="list-style-type: none"> <li>One sterile flocked vaginal swab for specimen collection</li> <li>One plastic vial with solvent solution. Solution contains 0.9% NaCl, 0.05% NaN<sub>3</sub> and 0.01% Triton X-100</li> <li>One PartoSure test strip in foil pouch with desiccant</li> </ul>	<ul style="list-style-type: none"> <li>Rapid fFN 10Q Cassette Kit</li> <li>PeriLynx™ (Hologic, Inc., Marlborough, MA, USA) analyser, printer, user manual and PeriLynx QCette® (Hologic, Inc., Marlborough, MA, USA) or Rapid fFN 10Q analyser, printer, user manual, and Rapid fFN 10Q QCette</li> <li>Rapid fFN Control Kit (Hologic, Inc., Marlborough, MA, USA)</li> <li>200-µl pipette</li> </ul>
Cost	£15 per test excluding VAT	£32 per test excluding VAT	£35 per test excluding VAT
Storage	The kit should be stored between 2 °C and 25 °C	The kit should be stored in a dry place between 4 °C and 25 °C	<p>The kit should be stored at room temperature between 15 °C and 30 °C</p> <p>Transport specimens at 2 °C to 25 °C, or frozen. Specimens are stable for up to 8 hours at room temperature</p> <p>Specimens not tested within 8 hours of collection must be stored refrigerated at 2 °C to 8 °C 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</p>
Test range	The test has a limit of detection of 10 µg/l and a measuring range of 10 to 8000 µg/l	The test has a limit of detection of 1 ng/ml and a measuring range of 1–40,000 ng/ml	The test has a detection range of 0–500 ng/ml, concentrations of > 500 ng/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 health-care professionals	The test is intended to be used in conjunction with other clinical information
VAT, value-added tax.			



### **Anticipated costs associated with the intervention**

The cost of the Rapid fFN 10Q System is usually £35 per test, not including value-added tax (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 (request for information from NICE to Hologic, Inc., 2017, personal communication from NICE).

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

The cost per PartoSure test (PAMG-1) is £32, not including VAT. No other costs are associated with this test (request for information from NICE to Parsagen Diagnostics Inc., 2017, personal communication from Parsagen Diagnostics Inc.).

## **Comparators**

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

### **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™ (Hologic, Inc., Marlborough, MA, USA) and Rapid fFN for the TLI<sub>Q</sub>® System (Hologic, Inc., Marlborough, MA, USA).<sup>29</sup>

#### **QuikCheck fFN**

QuikCheck fFN is a CE-marked, lateral flow immunoassay. The test kit includes a sterile applicator, test strip and a tube containing an extraction buffer. Additional materials required are a test tube rack and timer.<sup>29</sup>

The 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, the 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 indicate a positive result and a high risk of preterm delivery within 7–14 days; one line indicates a negative result and a low risk of delivery within 7–14 days; if no lines appear, the test result is invalid. The detection limit of the test is 50 ng/ml.<sup>29</sup>

The QuikCheck fFN test must be run within 15 minutes of the sample collection. The sample should be obtained before digital examination is conducted as cervix disruption may affect the test results. The presence of semen and gross vaginal bleeding may also affect the test results. The test is indicated for women presenting with threatened preterm labour and intact amniotic membranes.<sup>29</sup>

#### **Rapid Fetal Fibronectin for the TLI<sub>Q</sub> System**

Rapid fFN for the TLI<sub>Q</sub> System (Hologic, Inc., Marlborough, MA, USA) is a CE-marked immunochromatographic assay. The Rapid fFN test kit includes cassettes and a directional insert. Other materials required include a 200-µL pipette, a Rapid fFN Control Kit (includes positive control, negative control and directional insert) and the TLI<sub>Q</sub> System, which contains an analyser, printer and TLI<sub>Q</sub> QCette.<sup>29</sup>

A cervicovaginal sample is obtained from the posterior fornix or the ectocervical region of the external cervical os using a swab. The swab is rolled against the inside of the specimen transport tube to express the liquid into the extraction buffer and the swab is then discarded. The TLI<sub>Q</sub> analyser is set to internal incubation mode and the cassette containing the sample is inserted. Thereafter, 200 µl of the patient sample is dispensed into the sample application well of the Rapid fFN Cassette. After 20 minutes, the TLI<sub>Q</sub> analyser will display a result: positive, negative or invalid. The detection limit of the test is 50 ng/ml.<sup>29</sup>

This test is indicated for use in routine prenatal visits between 22<sup>+0</sup> weeks' and 30<sup>+6</sup> weeks' gestation, to assess the risk of delivery at  $\leq 7$  or  $\leq 14$  days from testing. Disruption to the cervix (i.e. through sexual intercourse, digital vaginal examination or transvaginal ultrasound scan) may result in a false-positive result. Douches, semen, white blood cells, red blood cells, bacteria and bilirubin may interfere with test results. However, if the patient reports sexual intercourse within the previous 24 hours, a negative fFN test result is still valid.<sup>29</sup>

### **Clinical assessment of symptoms alone**

Clinical assessment consists of taking a clinical history, observations of the woman and unborn baby and a speculum examination. See *Current guidelines* for more details.

## **Care pathways**

Clinical assessment and use of the biomarker tests aid clinicians in their decisions regarding 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.<sup>24</sup> 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 the 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.

Once a woman has been diagnosed with threatened preterm labour, she will typically be offered tocolytic therapy, corticosteroids and magnesium sulphate.<sup>24</sup>

### **Tocolytic therapy**

Tocolytic therapies increase the 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 (ACs).<sup>30</sup> There are many classes of tocolytic drugs with different mechanisms of action.<sup>30</sup>

The NICE guidelines<sup>24</sup> recommend nifedipine for women between 24<sup>+0</sup> and 33<sup>+6</sup> weeks' gestation in suspected or diagnosed labour with intact membranes. If nifedipine is contraindicated, NICE recommends oxytocin receptor antagonists [e.g. Atosiban (Sun Pharmaceuticals UK Ltd)] for tocolytic therapy.

Our clinical advisors suggest that tocolytic therapy is not commonly used in routine clinical practice. This may be attributable to recent evidence on the potential harms to the fetus and infant.<sup>30,31</sup>

### **Antenatal corticosteroids**

Antenatal corticosteroids [e.g. dexamethasone [Alliance Healthcare (Distribution) Ltd] or betamethasone (AAH Pharmaceuticals Ltd)] are prescribed in cases of threatened preterm labour to stimulate fetal lung development to reduce infant mortality and morbidity.<sup>32</sup> Following administration of steroids, there is a window within which the steroids appear to be most beneficial for the infant (*Figure 1C*, Norman *et al.*).<sup>33</sup> The primary documented negative effect of giving steroids is a negative reduction in birthweight of approximately 100 g.<sup>34</sup>

The National Institute for Health and Care Excellence currently recommends ACs for women between 26<sup>+0</sup> and 33<sup>+6</sup> weeks' gestation in suspected, diagnosed or established preterm birth, PPRM or undergoing planned preterm delivery. ACs should be considered for extremely preterm (between 24<sup>+0</sup> and 25<sup>+6</sup> weeks' gestation) and near-term (34<sup>+0</sup> and 35<sup>+6</sup> weeks' gestation) women in suspected, diagnosed and established preterm labour, PPRM or undergoing iatrogenic deliveries.<sup>24</sup> However, evidence regarding the

effectiveness of corticosteroid treatment at very low gestational ages remains uncertain.<sup>24</sup> NICE recommends that clinicians should discuss the benefits and risks associated with ACSs with the patient and their family. Repeat doses are contentious owing to possible risks, although this needs to be weighed against potential benefits of reduced RDS and serious adverse infant outcomes.<sup>32,35</sup> For this reason, some cases of repeat dosing may be acceptable depending on the interval since last dose, gestational age and likelihood of delivery within 48 hours.<sup>24</sup>

### Magnesium sulphate

Magnesium sulphate is a neuroprotective agent that significantly reduces neurological morbidities such as cerebral palsy in preterm infants.<sup>36</sup>

The National Institute for Health and Care Excellence recommends magnesium sulphate for women at 24<sup>+0</sup> to 29<sup>+6</sup> weeks' gestation in established labour or with iatrogenic delivery planned within 24 hours. Magnesium sulphate should also be considered for women between 30<sup>+0</sup> and 33<sup>+6</sup> weeks' gestation. A 4-g intravenous bolus dose of magnesium sulphate should be administered over 15 to 20 minutes, followed by an intravenous infusion of 1 g per hour for 24 hours or until delivery. Patients should be routinely monitored for signs of magnesium toxicity.

### Outcomes

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

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

$$\text{Sensitivity} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \quad (1)$$

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

$$\text{Specificity} = \frac{\text{True negative}}{\text{False positive} + \text{True negative}} \quad (2)$$

- Likelihood ratio (LR) – the likelihood of a given test result in a patient who has a preterm delivery compared with the likelihood of that same result in a patient who does not deliver preterm.

- Likelihood ratio for a positive test result (LR+) – how much more often a positive test result occurs in people who do deliver preterm compared with those who do not:

$$\text{Positive LR} = \frac{P(\text{Test + ve} \mid \text{preterm})}{P(\text{Test + ve} \mid \text{not preterm})} = \frac{\text{Sensitivity}}{1 - \text{Specificity}} \quad (3)$$

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

$$\text{Negative LR} = \frac{P(\text{Test - ve} \mid \text{preterm})}{P(\text{Test - ve} \mid \text{not preterm})} = \frac{1 - \text{Sensitivity}}{\text{Specificity}} \quad (4)$$

- Positive predictive value (PPV) – the probability of someone with a positive result actually having a preterm delivery:

$$\text{PPV} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \quad (5)$$

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

$$\text{NPV} = \frac{\text{True negative}}{\text{True negative} + \text{False negative}} \quad (6)$$

- Diagnostic yield (also known as test positivity rate or apparent prevalence) – 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.

## Chapter 2 Assessment of test accuracy

This report contains reference to confidential information provided as part of the NICE appraisal process. This information has been removed from the report and the results, discussions and conclusions of the report do not include the confidential information. These sections are clearly marked in the report.

### Methods for reviewing test accuracy

The diagnostic accuracies of PartoSure, Actim Partus and Rapid fFN 10Q Cassette Kit [at thresholds other than 50 ng/ml, referred to in the remainder of this report as quantitative fFN (qfFN)] were assessed by 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.<sup>37</sup> The protocol was registered on PROSPERO (reference number CRD42017072696).

#### Methods of the systematic review

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

### Identification of studies

#### Study sources and searches

To identify studies, the following bibliographic databases were searched from inception until July 2017 (the search was conducted in 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), Cumulative Index to Nursing and Allied Health Literature (CINAHL) (via EBSCOhost); BioSciences Information Service and Web of Science (via Clarivate Analytics) and The Cochrane Library [Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials (CENTRAL), Database of Abstracts of Reviews of Effects (DARE), Health Technology Assessment (HTA) database and NHS Economic Evaluation Database (NHS EED) (all via Wiley Interface)]. The search strategies were developed by a senior information specialist (CC), and comprised 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<sup>38</sup> 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 [Clarivate Analytics (formerly Thomson Reuters), Philadelphia, PA, USA] and deduplicated using automatic and manual checking.

Additional sources were searched:

- 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 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 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 ClinicalTrials.gov website [<https://clinicaltrials.gov/ct2/home> (accessed July 2017)] and ISRCTN (International Standard Randomised Controlled Trial Number) [[www.isrctn.com/editAdvancedSearch](http://www.isrctn.com/editAdvancedSearch) (accessed July 2017)] using terms designed to identify the index tests (see *Appendix 1*).

- Google Advanced Search (Google Inc., Mountain View, CA, USA) was used to conduct web searching (September 2017), using terms designed to identify the index tests (see *Appendix 1*). For each term searched using Google Advanced Search, 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.

### **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 and HC) and screened for possible inclusion, using prespecified inclusion and exclusion criteria (see *Inclusion and exclusion criteria*). 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 and HC) independently examined full texts for inclusion or exclusion. Disagreements were again resolved by discussion within the review team.

### **Inclusion and exclusion criteria**

#### **Population**

In line with the NICE scope,<sup>12</sup> 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 at high or low risk for preterm labour. Studies were also eligible regardless of whether or not it was stipulated that the recruited population had access to transvaginal ultrasound scans. There were no specific inclusion criteria relating to the number of weeks of gestation of the women recruited; however, the study was required to define its population as preterm. The unit of assessment was individual women with a single result for each test.

Initially, studies were included only if all participants were expecting a singleton pregnancy. However, owing to the lack of evidence, a protocol amendment was made to include studies in which twin or multiple pregnancies were included but made up  $\leq 20\%$  of the total population recruited. We are not aware of any published evidence to suggest that multifetal pregnancies would alter the DTA of any of the tests.

#### **Index tests**

In accordance with the NICE scope,<sup>12</sup> 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 (qfFN), used with a threshold other than 50 ng/ml (with or without a clinical assessment).

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

#### **Reference standard**

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

- preterm delivery within 48 hours or within 7 days
- clinical assessment of symptoms alone
- fetal fibronectin at a threshold of 50 ng/ml (qualitative or quantitative test).

In addition, studies that provided test accuracy data by comparing the results of one index test with 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 48 hours, or within 7 days, as the reference standard.

### Outcomes

In accordance with the NICE scope,<sup>12</sup> 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
- likelihood ratio for negative test result
- positive predictive value – true positive/(true positive + false positive)
- negative predictive value – 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 48 hours
- test failure (non-informative test result) rate
- time to test result.

### 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 that data were reported specifically for all women receiving the index test. Two-gate diagnostic studies were also eligible for inclusion.

Studies in which 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 the window stipulated in the manufacturers' guidelines owing 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.

### 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.

### Critical appraisal strategy

The methodological quality of the studies was assessed by one reviewer (SD) and judgements were checked by a second reviewer (HC), in accordance with criteria specified by phase 3 of the QUADAS-2 tool (see *Appendix 2*).<sup>39</sup> Any disagreements were resolved by discussion, with involvement of a third reviewer (JVC) as necessary.

## Methods of data synthesis

For all included studies, sensitivity, specificity, positive and negative predictive value, positive and negative LR, 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 qfFN they 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,<sup>40</sup> 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.

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.*<sup>41</sup> No meta-analysis was undertaken for qfFN at any threshold or for the 48-hour reference standard for PartoSure.

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

## Results of the systematic review

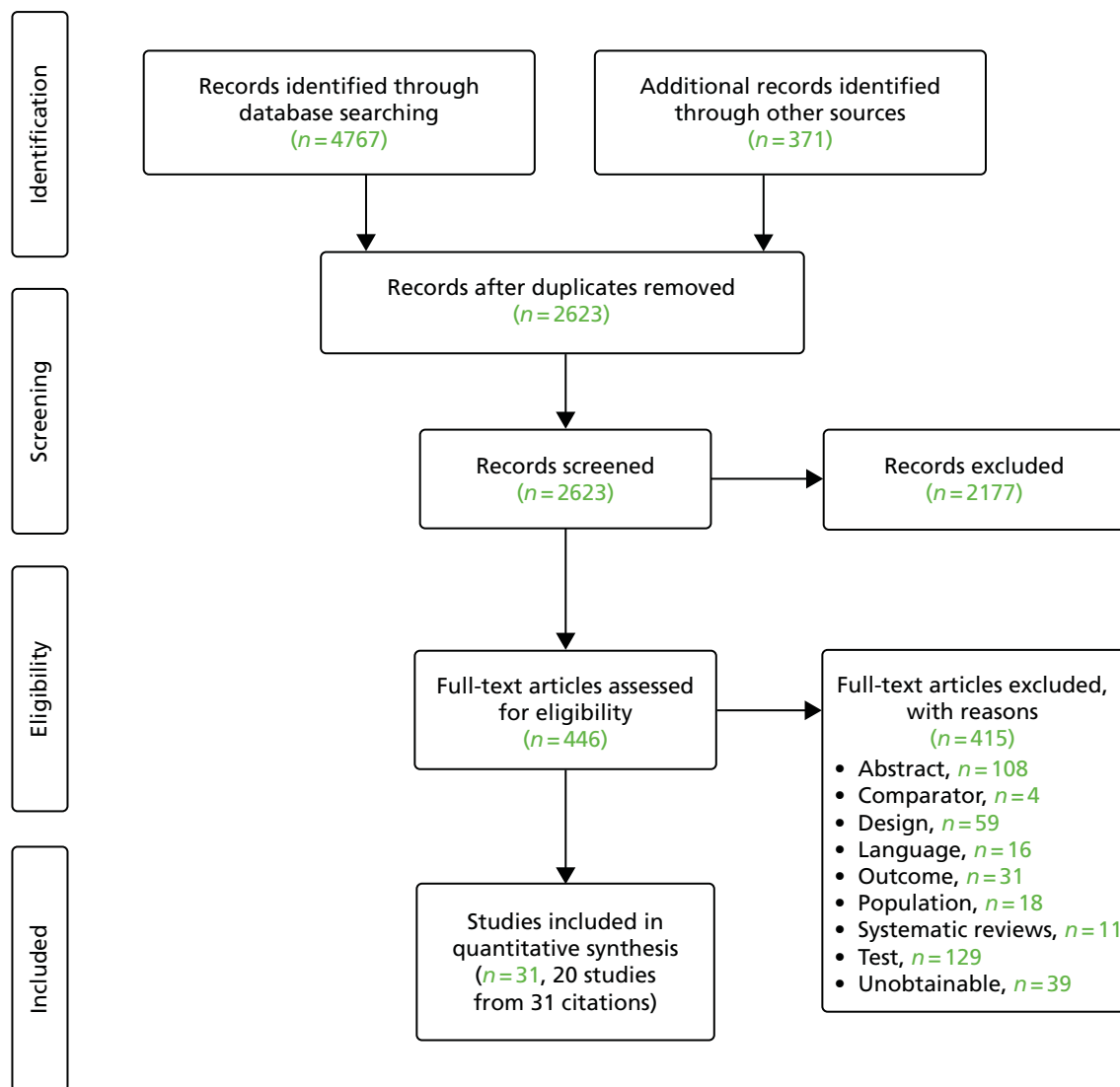
In this section, the results of the systematic review of PartoSure, Actim Partus and qfFN are presented. Studies providing DTA data for one or more of these index tests are included (see *Inclusion and exclusion criteria* for specific details on inclusion and exclusion criteria).

### Overview of the quantity and quality of research available

The searches retrieved a total of 2619 unique titles and abstracts. A total of 2177 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 an irrelevant test, typically qualitative fFN ( $n = 129$ ), the study design ( $n = 59$ ), that outcomes did not match the review inclusion criteria ( $n = 31$ ) or that the article was an abstract that both had 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 3*. Additional tables (see *Tables 28–30*) are provided in *Appendix 3*, listing all the citations provided by the industry to NICE along with whether or not the citation was included, and, if not, the reason for exclusion. After screening relevant systematic reviews ( $n = 11$ ; see *Appendix 3, Included systematic reviews*), 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 2*.





**FIGURE 2** Summary of the selection process.

### Ongoing trials

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

Four of the ongoing trials are based in the UK, two of which are planning to enrol > 1000 participants (ISRCTN41598423, n = 2100, and Integrated Research Application System ID 111142, n = 1181). However, it was not possible to include data from these ongoing trials in this review of test accuracy.

### Description of the included studies

Characteristics of the included studies are summarised in *Table 3*. Two studies, Assessment of Perinatal Outcome after Sustained Tocolysis in Early Labour (APOSTEL-1)<sup>42,43</sup> and Hadzi-Lega *et al.*,<sup>44</sup> assessed the DTA of two different index tests in the same population: APOSTEL-1<sup>42,43</sup> assessed both Actim Partus and qfFN whereas the study by Hadzi-Lega *et al.*<sup>44</sup> assessed Actim Partus and PartoSure. APOSTEL-1<sup>42,43</sup> was one of the larger studies (n = 350), conducted in 10 centres around the Netherlands, whereas the study by Hadzi-Lega *et al.*<sup>44</sup> was smaller (n = 57), from one centre in Macedonia.

A further 14 studies<sup>45–58</sup> assessed the DTA for Actim Partus only. Three studies<sup>41,59–61</sup> assessed PartoSure only and one<sup>62</sup> assessed qfFN only.

TABLE 2 Ongoing trials

Study	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	USA	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 (qfFN) help diagnosis and clinical decision making? (QUIDS and QUIDS-2)	HTA 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
(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)

IRAS, Integrated Research Application System.

TABLE 3 Study characteristics

Study (first author/study name and year)	Other tests used	Number included (number recruited)	Country (number of centres)	Definition of preterm labour symptoms	Weeks of gestation	Dilatation threshold for exclusion	Other exclusion criteria
<b>Study assessing Actim Partus and qfFN</b>							
APOSTEL-1 (2016) <sup>42,43</sup>	Cervical length	350 (714)	The Netherlands (10)	Uterine contractions (> 3/30 minutes), abdominal pain, back pain and vaginal bleeding	24–34	> 3 cm	Contraindications for tocolysis, iatrogenic deliveries and tocolytic treatment prior to testing
<b>Study assessing Actim Partus and PartoSure</b>							
Hadzi-Lega (2017) <sup>44</sup>	Cervical length	57 (72)	Macedonia (1)	Uterine contractions and abdominal pain	22–34 <sup>+6</sup>	> 3 cm	Antepartum haemorrhage, cervical cerclage and multiple gestations
<b>Actim Partus</b>							
Abo El-Ezz (2014) <sup>45</sup>	N/A	57 (80)	Kuwait (2)	Uterine contractions (≥ 8/hour), back pain, pelvic pressure, vaginal discharge and 50% effacement	24–34	> 3 cm	Cervical cerclage, chorioamnionitis, fetal abnormalities, intrauterine growth restriction, multiple gestations, placenta praevia, prior cervical examination, sexual intercourse in previous 24 hours, uterine anomalies and vaginal bleeding
Altinkaya (2009) <sup>46</sup>	N/A	105 (NR)	Turkey (1)	Uterine contractions	24–34	≥ 2 cm	Fetal abnormalities, history of preterm delivery, intrauterine growth restriction, multiple gestations, preeclampsia, smokers, uterine anomalies and vaginal bleeding
Azlin (2010) <sup>47</sup>	Cervical length	51 (51)	Malaysia (NR)	Uterine contractions	24–36	≥ 3 cm	Abruptio placenta, cervical cerclage, cervical incompetence, multiple gestations and placenta praevia
Brik (2010) <sup>48</sup>	N/A	276 (325)	Spain (1)	Uterine contractions, abdominal pain, back pain, leaking of fluid and other	24–34	> 3 cm	Abruptio placenta, cervical cerclage, fetal abnormalities, fetal distress and vaginal bleeding (active labour)

continued

TABLE 3 Study characteristics (continued)

Study (first author/study name and year)	Other tests used	Number included (number recruited)	Country (number of centres)	Definition of preterm labour symptoms	Weeks of gestation	Dilatation threshold for exclusion	Other exclusion criteria
Cooper (2012) <sup>49</sup>	Qualitative fFN (unclear which test)	349 (366)	Canada (2)	Symptoms judged by physician to be indicative of labour	24 <sup>+0</sup> to 34 <sup>+6</sup>	NR	Antepartum haemorrhage and chorioamnionitis (active labour)
Danti (2011) <sup>50</sup>	Cervical length	60 (102)	Italy (1)	Uterine contractions ( $\geq 4/20$ minutes)	24 <sup>+0</sup> to 32 <sup>+6</sup>	> 3 cm	Abruptio placenta, cervical cerclage, fetal abnormalities, intrauterine growth restriction, multiple gestations, placenta praevia, preeclampsia, uterine anomalies and vaginal bleeding
Eroglu (2007) <sup>51</sup>	QuikCheck fFN, cervical length	51 (51)	Turkey (1)	Uterine contractions (> 10/hour)	24–35	$\geq 3$ cm	Abruptio placenta, fetal abnormalities, intrauterine growth restriction, multiple gestations, placenta praevia, preeclampsia, sexual intercourse in previous 24 hours, uterine anomalies and vaginal bleeding
Goyal (2016) <sup>52</sup>	Cervical length	60 (95)	India (1)	Uterine contractions (> 4/20 minutes) and abdominal pain	24–36	NR	Fetal abnormalities, fetal growth restrictions, preeclampsia, multiple gestations and vaginal bleeding
Lembet (2002) <sup>53</sup>	N/A	36 (36)	Turkey (1)	Uterine contractions (> 10/hour)	20–36	N/A	Fetal abnormality, intrauterine growth restriction, preeclampsia, multiple gestations, uterine anomalies and vaginal bleeding
Riboni (2011) <sup>54</sup>	fFN by ELISA	210	Italy (2)	Uterine contractions (> 10/hour)	24–34	> 2 cm	Fetal abnormalities, multiple gestations, placenta praevia, prior cervical examination, sexual intercourse in previous 24 hours, uterine anomalies and vaginal bleeding
Tanir (2009) <sup>55</sup>	N/A	68 (121)	Turkey (1)	Uterine contractions (> 4/20 minutes), changes in cervix, back pain and increased discharge	24–37	$\geq 3$ cm	Asthma, cervical cerclage, diabetes mellitus, digital examination in previous 24 hours, hyperthyroidism, multiple gestations, preeclampsia, sexual intercourse in previous 24 hours, tocolytic treatment prior to testing, vaginal bleeding and vaginal douche in previous 24 hours

Study (first author/study name and year)	Other tests used	Number included (number recruited)	Country (number of centres)	Definition of preterm labour symptoms	Weeks of gestation	Dilatation threshold for exclusion	Other exclusion criteria
Ting (2007) <sup>56</sup>	Qualitative fFN (unclear which test)	94 (108)	Singapore (1)	NR	24–34	≥ 3 cm	Cervical cerclage, chorioamnionitis, fetal asphyxia, fetal abnormalities, intrauterine growth restrictions, multiple gestations, placenta praevia and preeclampsia
Tripathi (2016) <sup>57</sup>	QuikCheck fFN	468 (550)	India (1)	Uterine contractions (> 1/10 minutes) and labour pains	28 <sup>+1</sup> to 36 <sup>+6</sup>	> 3 cm	Blood-mixed cervical secretions, diarrhoea, prepartum haemorrhage, previous preterm delivery, sexual intercourse in previous 24 hours, urinary tract infection and vaginal leakage
Vishwekar (2017) <sup>58</sup>	N/A	30 (NR)	India (1)	Uterine contractions and vaginal discharge	28–37	NR	Blood-mixed cervical secretions, fetal distress, hypertension and intrauterine growth restrictions (active labour)
<b>PartoSure</b>							
Bolotskikh (2017) <sup>59</sup>	Cervical length	99 (100)	Russia (1)	Uterine contractions, abdominal pain, back pain, pelvic pressure, menstrual-like cramping and diarrhoea	22 <sup>+0</sup> to 36 <sup>+6</sup>	> 3 cm	Maternal age < 18 years, multiple gestations, prior cervical examination, placenta praevia, symptoms unrelated to threatened preterm delivery (e.g. trauma), tocolytic treatment prior to testing and vaginal bleeding
Nikolova (2014/15) <sup>60,61</sup>	Cervical length, QuikCheck fFN	203 (219)	Macedonia and Russia (2)	Uterine contractions, abdominal pain and pelvic pressure	20 <sup>+0</sup> to 36 <sup>+6</sup>	> 3 cm	Cervical cerclage, placenta praevia, maternal age < 18 years and multiple gestations
Werlen (2015) <sup>41</sup>	N/A	41 (42)	France (1)	Uterine contractions and cervical changes	24–34	> 3 cm	Blood-mixed cervical secretions, multiple gestations and vaginal infection
<b>qfFN</b>							
Bruijn (2016) <sup>62</sup>	Cervical length	455 (484)	The Netherlands, Switzerland, Belgium, Germany and Austria (10)	Uterine contractions (> 3/30 minutes), abdominal pain, back pain and vaginal bleeding	24–34	> 3 cm	Contraindications for tocolysis, fetal distress, iatrogenic deliveries, tocolytic treatment prior to testing and triplet or higher gestations
ELISA, enzyme-linked immunosorbent assay; N/A, not applicable; NR, not reported.							

For Actim Partus, study sizes ranged from 30 in Vishwekar *et al.*<sup>58</sup> to 468 in Tripathi *et al.*<sup>57</sup> and covered the following countries: Kuwait,<sup>45</sup> Turkey,<sup>46,51,53,55</sup> Malaysia,<sup>47</sup> Spain,<sup>48</sup> Canada,<sup>49</sup> Italy,<sup>50,54</sup> India<sup>52,57,58</sup> and Singapore.<sup>56</sup> The three studies assessing PartoSure were conducted in Russia,<sup>59</sup> Macedonia<sup>60,61</sup> and France<sup>41</sup> and the study size ranged from 41 in Werlen *et al.*<sup>41</sup> to 203 in Nikolova *et al.*<sup>61</sup> Finally, the Bruijn study<sup>62</sup> assessed qfFN only and recruited 455 participants from 10 centres across Austria, Belgium, Germany, the Netherlands and Switzerland.

### Key differences between studies

It was notable that the prevalence rates of preterm birth differed greatly between studies (see *Tables 6 and 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 of gestation) differed between studies. The inclusion/exclusion criteria of the studies also differed regarding the presenting symptoms of the women, the proportion of women with singleton gestations, the risk status of included women, dilatation thresholds applied and other specific exclusion criteria.

### *Differences between studies in prevalence of preterm birth*

There was clear variation between studies regarding the prevalence of the reference standard (i.e. the prevalence of preterm birth within 7 days and within 48 hours). Across all 20 studies, prevalence of preterm birth within 7 days ranged from 1.7% [95% confidence interval (CI) 0.6% to 3.7%] in Cooper *et al.*<sup>49</sup> to 73.3% (95% CI 60.3% to 83.9%) in Goyal *et al.*<sup>52</sup> Both of these studies assessed the test accuracy of Actim Partus. Therefore, the studies assessing Actim Partus 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 within 7 days still ranged widely (from 2.4%, 95% CI 0.1% to 12.9% in Werlen *et al.*<sup>41</sup> to 17.2%, 95% CI 12.3% to 23.2% in Nikolova *et al.*<sup>61</sup>). In the studies assessing qfFN, prevalence of preterm birth within 7 days was slightly higher in APOSTEL-1<sup>42,43</sup> than in EUIFS<sup>62</sup> (19.7%, 95% CI 15.7% to 24.3% vs. 10.5%, 95% CI 7.9% to 13.7%, respectively).

Seven studies provided DTA data for the index tests against preterm birth within 48 hours.<sup>41,48,52,53,56–58</sup> Across these seven studies, the prevalence of preterm birth within 48 hours ranged from 2.4% (95% CI 0.1% to 12.9%) in Werlen *et al.*<sup>41</sup> to 58.3% (95% CI 44.9% to 70.9%) in Goyal *et al.*<sup>52</sup> The study by Werlen *et al.*<sup>41</sup> was the only study assessing PartoSure against preterm birth within 48 hours, with the other six studies<sup>48,52,53,56–58</sup> assessing Actim Partus. The lowest prevalence of preterm birth within 48 hours within these six Actim Partus studies was 5.3% (95% CI 1.7% to 12.0%) in Ting *et al.*<sup>56</sup>

These differences in prevalence displayed between studies are probably attributable to differences in the populations recruited into the studies (e.g. differences in gestational age and in presenting symptoms of preterm labour; see *Differences between studies in inclusion/exclusion criteria*) and will also probably have an impact on the DTA data presented in *Results of quantitative data synthesis (test accuracy data)* and the generalisability of these data to the NHS in England.

### *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 takes place within this time frame, it remains unclear whether or not a spontaneous delivery may have occurred, which thus makes it impossible to accurately assess the reference standard in these women. Nine of the included studies<sup>41,45,46,50–52,54,56,57</sup> did not report the mode of delivery (i.e. whether or not birth was spontaneous, or whether or not there were any planned caesarean sections or inductions).

The other 11 studies<sup>42–44,47–49,53,55,58–62</sup> provided some data regarding mode of delivery (*Table 4*). Four of these studies (APOSTEL-1,<sup>42,43</sup> Hadzi-Lega *et al.*,<sup>44</sup> EUIFS<sup>62</sup> and Bolotskikh *et al.*<sup>59</sup>) reported that women who had a non-spontaneous delivery within the time frame of the reference standard were excluded from the

TABLE 4 Participant characteristics

Study (first author/study name and year)	Participants (n)	Maternal age (years), mean (SD) [range]	Gestational age at presentation (weeks), mean (SD) [range]	Multiple gestations, n (%)	BMI (kg/m <sup>2</sup> ), mean (SD)	Gravidity, mean (SD)	Parity, mean (SD)	Previous preterm delivery, n (%)	Previous miscarriage/stillbirth, n (%)	Mode of delivery, n (%)
<b>Study assessing Actim Partus and qffN</b>										
APOSTEL-1 <sup>42,43</sup>	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 time frame excluded
<b>Study assessing Actim Partus and PartoSure</b>										
Hadzi-Lega (2017) <sup>44</sup>	57	Median 27 [IQR 23.0–30.5]	Median 31 [IQR 28.8–32.4]	0 (0)	NR	NR	NR	NR	NR	Non-spontaneous deliveries within reference-standard time frame excluded
<b>Actim Partus</b>										
Abo El-Ezz (2014) <sup>45</sup>	57	27.40 (6.1)	29.70 (2.5)	0 (0)	NR	NR	2.91 (NR)	NR	NR	NR
Altinkaya (2009) <sup>46</sup>	105 <sup>a</sup>	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) <sup>47</sup>										
phiGFBP-1 (+)	7	29.57 (3.99)	32.96 (3.07) <sup>b</sup>	0 (0)	NR	2.43 (1.27)	1.00 (1.16)	NR	0.43 (1.13)	<ul style="list-style-type: none"> <li>Spontaneous: 36 (70.6)</li> <li>Elective caesarean section: 7 (13.7)</li> <li>Emergency caesarean section: 4 (7.8)</li> <li>Unknown: 4 (7.8)</li> </ul>
phiGFBP-1 (-)	44	28.34 (4.32)	32.38 (2.64) <sup>b</sup>	0 (0)	NR	2.59 (1.59)	0.91 (0.96)	NR	0.68 (1.25)	
Brik (2010) <sup>48</sup>	276	29.4 (5.9) [15–46]	29.9 (2.8) [23–34]	0 (0)	NR	NR	Nulliparous = 58.3%	26 (9.4)	NR	Unclear
Cooper (2012) <sup>49</sup>	349	29 (5.0) [17–46]	Median 29 <sup>+6</sup> (4 <sup>+6</sup> ) [IQR 24–34]	20 (5.7) <sup>c</sup>	NR	NR	Nulliparous = 43.3%	56 (16.1)	NR	<ul style="list-style-type: none"> <li>Spontaneous: 182 (52.1)</li> <li>Operative: 52 (14.9)</li> <li>Caesarean section: 115 (33.0)</li> </ul>
Danti (2011) <sup>50</sup>	60 <sup>d</sup>	Median 31 [IQR 28–34]	Median 30.0 [IQR 28.7–31.4]	0 (0)	NR	NR	Nulliparous = 63%	NR	NR	NR
Eroglu (2007) <sup>51</sup>	51 <sup>a</sup>	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) <sup>52</sup>	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) <sup>53</sup>	36 <sup>a</sup>	28.4 (5.3)	31.3 (3.3)	0 (0)	< 19.6, n = 7; > 26, n = 29	2.2 (1.6)	0.7 (1.1)	7 (16)	3 (7)	Unclear
Riboni (2011) <sup>54</sup>	210		28.7 (NR)	0 (0)		NR	NR	NR	NR	NR
Delivering term		30.4 (5.6)			23.36 (3.99)					
Delivering preterm		30.7 (5.1)			22.72 (3.7)					

continued

TABLE 4 Participant characteristics (continued)

Study (first author/study name and year)	Participants (n)	Maternal age (years), mean (SD) [range]	Gestational age at presentation (weeks), mean (SD) [range]	Multiple gestations, n (%)	BMI (kg/m <sup>2</sup> ), mean (SD)	Gravidity, mean (SD)	Parity, mean (SD)	Previous preterm delivery, n (%)	Previous miscarriage/stillbirth, n (%)	Mode of delivery, n (%)
<b>Tanir (2009)<sup>55</sup></b>										
phiGFBP-1 (+) <sup>e</sup>	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)	<ul style="list-style-type: none"> <li>Caesarean section: 18 (72.0)<sup>f</sup></li> <li>Vaginal: 16 (64)<sup>g</sup></li> </ul>
phiGFBP-1 (-) <sup>g</sup>	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)	<ul style="list-style-type: none"> <li>Caesarean section: 20 (46)<sup>f</sup></li> <li>Vaginal: 14 (33)<sup>g</sup></li> </ul>
<b>Ting (2007)<sup>56</sup></b>										
phiGFBP-1 (+) <sup>e</sup>	28	27	30.5	0 (0)	NR	2	0.5	NR	NR	NR
phiGFBP-1 (-) <sup>g</sup>	66	27	32.2	0 (0)	NR	2	1	NR	NR	NR
Tripathi, 2016 <sup>57</sup>	468	NR	NR	NR	NR	NR	NR	0 (0)	NR	NR
Vishwekar 2017 <sup>58</sup>		25 [19–35]		2 (6.7)	'Normal limits'	NR	NR	4 (13.3)	NR	Unclear
phiGFBP-1 (+)	14		32							
phiGFBP-1 (-)	16		32.5							
<b>PartoSure</b>										
Bolotskikh (2017) <sup>59</sup>	99 (100)	Median 25 [IQR 23–38]	Median 32 [IQR 29–36]	0 (0)	NR	NR	Nulliparous = 32%	15 (15)	27 (27)	Non-spontaneous deliveries within reference-standard time frame excluded
Nikolova (2014 and 2015) <sup>60,61</sup>	203	Median 27 [range 18–43]	Median 32.0 [range 20.5–36.6]	0 (0)	NR	NR	NR	NR	NR	Patients with non-spontaneous delivery excluded (n = 8); unclear when these happened
Werlen (2015) <sup>41</sup>	41	27.6 (5.3) [18–39]	29.5 (2.91) [24–34]	0 (0)	NR	NR	0.54 (0.71)	NR	NR	NR
<b>qfFN</b>										
EUFS <sup>62</sup>	455	29.5 (5.2)	Median 29.6 [IQR 26.7–31.6]	67 (15)	Median 24.5 (IQR 22.0–28.0) (n = 429)	NR	Nulliparous = 55%	72 (16)	NR	Non-spontaneous deliveries within reference-standard time frame excluded

IQR, interquartile range; NR, not reported; SD, standard deviation.

a Study group (symptomatic).

b Reported in paper as POA, assumed to mean presentation on admission for gestational weeks.

c Two participants; number of fetuses unknown.

d High-risk group: cervical length ≤ 30 mm.

e Actim Partus test positive.

f Data as reported in paper; however, values do not add up.

g Actim Partus test negative.



test accuracy data. Women who had a non-spontaneous delivery outside 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 48 hours or within 7 days). In a further three studies (Vishwekar *et al.*,<sup>58</sup> Brik *et al.*<sup>48</sup> and Nikolova *et al.*<sup>60,61</sup>), iatrogenic delivery was mentioned as a reason for exclusion, but it is unclear how many of these deliveries took place within the time frame of the reference standard, and in another study (Lembet *et al.*<sup>53</sup>) the number of iatrogenic deliveries could not be ascertained.

In three studies,<sup>47,49,55</sup> the numbers of spontaneous/iatrogenic deliveries were reported, but no exclusion of data from non-spontaneous deliveries was made. In the study by Azlin *et al.*,<sup>47</sup> 70.6% of women delivered spontaneously, 7.8% underwent an emergency caesarean section and for a further 7.8% there were no data available on mode of delivery. Although 13.7% of women delivered by a planned caesarean section, it is unclear whether or not these took place within the time frame of the reference standard, and these women were not excluded from the test accuracy data.<sup>47</sup> In Cooper *et al.*,<sup>49</sup> 52.1% of women delivered spontaneously, 14.9% had an operative delivery and 33.0% had a caesarean section, although, again, it was unclear how many of these procedures were planned and how many took place within the time frame of the reference standard. Finally, Tanir *et al.*<sup>55</sup> report the proportion of women who delivered by caesarean section or from vaginal delivery; however, these data do not appear to be correct (see *Table 4*).

### **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 had an impact on the 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 onwards (see *Table 3*).<sup>41–43,45–52,54–56,62</sup> In these 14 studies, the upper limit of gestation varied widely: eight studies recruited women at gestation of up to 34 weeks,<sup>41–43,45,46,48,54,56,62</sup> with the upper limit of gestation ranging from 32<sup>+6</sup> weeks (Danti *et al.*<sup>50</sup>) to 37 weeks (Tanir *et al.*<sup>55</sup>) in the remaining six studies.

Of the remaining six studies, four recruited women at an earlier gestation: two studies<sup>53,61</sup> recruited women from as early as 20 weeks' gestation and two studies<sup>44,59</sup> included women from 22 weeks' gestation. Two of these studies recruited women up until 36<sup>+6</sup> weeks,<sup>59,61</sup> one recruited women up until 36 weeks<sup>53</sup> and the other recruited women up until 34<sup>+6</sup> weeks.<sup>44</sup> The other two studies<sup>57,58</sup> recruited women at a later gestation (i.e. from 28 weeks' gestation), with both recruiting up until 37 weeks (36<sup>+6</sup> weeks for Tripathi *et al.*<sup>57</sup>).

None of the studies presented test accuracy data between different gestational cut-off points. It was not possible, therefore, to make any within-study assessment, for any of the index tests, whether or not test accuracy differed based on gestation.<sup>41–62</sup>

**Presenting symptoms of preterm labour** Other than stating that women had to be symptomatic, all studies except for Ting *et al.*<sup>56</sup> provided some further details about the presenting symptoms of preterm labour. However, one further study (Cooper *et al.*<sup>49</sup>) added only that that 'symptoms indicative of labour were to be determined by a physician'.

All other studies reported uterine contractions as a necessary indicator of preterm labour.<sup>41–48,50–55,57–62</sup> Ten studies additionally described the rate of uterine contractions necessary for inclusion: a rate of six contractions per hour was reported by Tripathi *et al.*,<sup>57</sup> Buijn<sup>62</sup> and APOSTEL-1,<sup>42,43</sup> a rate of eight contractions per hour was reported by Abo El-Ezz *et al.*,<sup>45</sup> a rate of 10 contractions per hour was reported by Eroglu *et al.*,<sup>51</sup> Riboni *et al.*<sup>54</sup> and Lembet *et al.*<sup>53</sup> and a rate of 12 contractions per hour was reported by Danti *et al.*,<sup>50</sup> Goyal *et al.*<sup>52</sup> and Tanir *et al.*<sup>55</sup>

Other commonly reported symptoms included in definitions of preterm labour were abdominal pain (in seven studies<sup>42–44,48,52,59–62</sup>), back pain (in six studies<sup>42,43,45,48,55,59,62</sup>), pelvic pressure (in three studies<sup>45,59–61</sup>), vaginal bleeding (in the APOSTEL-1 and Buijn studies<sup>42,43,62</sup>) and vaginal discharge (in three studies<sup>45,55,58</sup>).

**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 (qfFN and Actim Partus)] and 20% of the included pregnancies were multiple pregnancies.<sup>42,43</sup> Two of these studies assessed only Actim Partus, with one (Cooper *et al.*<sup>49</sup>) having 6% of included pregnancies as multiple pregnancies and the other (Vishwekar *et al.*<sup>58</sup>) having 7% of included pregnancies as multiple pregnancies. The final study that included multiple pregnancies (Bruijn<sup>62</sup>) only assessed qfFN, and they made up 15% of the population. We are not aware of any published evidence to suggest that multifetal pregnancies would alter the DTA of either qfFN or Actim Partus.

**Risk status of participants** Only one of the included studies (Bolotskikh *et al.*<sup>59</sup>) clearly reported the risk status of included women (i.e. whether or not the women were at high or low risk for preterm labour prior to the onset of symptoms). In this study, the population was described as high risk because 15% of the participants had previously experienced preterm labour, 43% had mild preeclampsia and 51% were previously hospitalised during the pregnancy.<sup>59</sup> However, although none of the other studies explicitly stated that the populations were at high risk of preterm labour, eight additional studies<sup>42,43,48,49,51–53,58,62</sup> also recruited some women who had previously experienced preterm delivery (see *Table 4*). In addition, two studies (APOSTEL-1<sup>42,43</sup> and Danti *et al.*<sup>50</sup>) restricted their population by conducting tests in women presenting with a cervical length of < 30 mm. However, this high-risk status is associated with symptoms at presentation rather than the women being high risk prior to the onset of symptoms.

Recruiting high-risk women would be expected to have an impact on the prevalence of preterm birth. However, because almost all studies did not clearly report risk status, it is not possible to properly assess if or how this had an impact on prevalence rates (and, therefore, test accuracy data).

**Dilatation threshold and cervical length** All studies except Cooper *et al.*,<sup>49</sup> Goyal *et al.*,<sup>52</sup> Lembet *et al.*<sup>53</sup> and Vishwekar *et al.*<sup>58</sup> included a dilatation threshold for exclusion; typically, the threshold was > 3 cm or  $\geq$  3 cm. However, Riboni *et al.*<sup>54</sup> had a dilatation threshold of > 2 cm and Altinkaya *et al.*<sup>46</sup> had a threshold of  $\geq$  2 cm.

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.<sup>42–44,47,50–52,59–62</sup> In seven of these studies,<sup>44,47,51,52,59–62</sup> no selection of women took place in accordance with 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. However, in the other two studies (APOSTEL-1<sup>42,43</sup> and Danti *et al.*<sup>50</sup>), all women included in final analyses of index test data had a transvaginal cervical length measurement of  $\leq$  30 mm, which would probably increase the prevalence of preterm birth in these studies. Both of these studies were assessing Actim Partus, with APOSTEL-1 additionally assessing qfFN.<sup>42,43,50</sup>

**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 24 hours, fetal asphyxia, fetal abnormalities, fetal distress, history of preterm delivery, hypertension, hyperthyroidism, iatrogenic deliveries, intrauterine growth restriction, maternal age of < 18 years, multiple gestations, placenta praevia, preeclampsia, prior cervical examination, sexual intercourse in the previous 24 hours, 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 24 hours, vaginal infection and vaginal leakage.<sup>41–62</sup> All exclusion criteria were reasonable in the context of the index tests under consideration (see *Quality appraisal summary*).

## Summary of the reference standard

In all studies, the reference standard was preterm birth, within 48 hours and/or within 7 days.<sup>41–62</sup> All 20 included studies evaluated the index tests against the 7-day reference standard.<sup>41–62</sup> Six of the Actim Partus studies<sup>48,52,53,56–58</sup> and one PartoSure study<sup>41</sup> also evaluated the index test against a 48-hour reference standard. qfFN was not evaluated against the 48-hour reference standard.<sup>42,43,62</sup>

## Summary of test administration

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

### *Actim Partus*

In the 16 studies that used the Actim Partus test,<sup>43,45–58</sup> the information provided on how the test was administered typically followed the manufacturer's guidance; however, there were some differences. The reporting of detection-limit thresholds varied between studies, but it is likely that all studies used a threshold of 10 µg/l: eight studies clearly reported a detection limit of 10 µg/ml.<sup>45,49,50,52–54,57,58</sup> Two studies<sup>46,51</sup> reported that samples higher than 30 µg/l give 'a strong positive result'. It is unclear in both of these studies whether or not a weak positive at 10 µg/l would have been considered a positive result, although this appears to be the case.<sup>46,51</sup> One study<sup>48</sup> states that a threshold of 30 µg/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 10 µg/l). Finally, five studies<sup>41,43,44,47,55,56</sup> did not report what detection limit they used; however, given the qualitative nature of the test, it appears most likely that a 10-µg/l threshold was used. Indeed, the manufacturer's guidance indicates that a concentration of ≥ 10 µg/l in the cervical fluid causes a positive Actim Partus test reaction result.

All studies report taking their sample from around the external cervical orifice or cervical os or as a cervical specimen.<sup>43–58</sup> However, two studies<sup>43,56</sup> report that the sample was taken from the posterior fornix. 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 the posterior fornix would probably yield a higher false-negative rate because 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 *et al.*,<sup>63</sup> although using the qfFN test, it is likely that Actim Partus would be affected in a similar manner.

The manufacturer's instructions state that if the single (control) line does not appear then the test is invalid. However, one study<sup>55</sup> interpreted no visible lines as a positive test result. This was their way of dealing with missing data owing to invalid results. This is unlikely to greatly alter results as only two tests from 68 results were invalid.<sup>55</sup> The remaining studies did not report details on how invalid tests were treated.

Two studies<sup>43,49</sup> both froze their samples at –20 °C for future analysis. In APOSTEL-1,<sup>43</sup> samples were reported to have been transferred and stored at –80 °C within 6 months. 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.<sup>43</sup> Likewise, in the study by Cooper *et al.*<sup>49</sup> it is unclear how long the tests remained frozen before testing. Both of these studies<sup>43,49</sup> go on to describe that samples were thawed before the Actim Partus test was run. This protocol differs from the manufacturer's guidance (and, of course, clinical practice), in which freezing a sample is not discussed in the instructions for use. We have received clinical input from our obstetricians to suggest that freezing is unlikely to affect the sample's integrity and, therefore, is unlikely to have an impact on test accuracy.

### *PartoSure*

All four studies assessing the PartoSure test appeared to conduct the test in a manner that was consistent with the manufacturer's guidance.<sup>41,44,59–61</sup> However, Bolotskikh *et al.*<sup>59</sup> did not specifically report that samples were collected using a speculum.

### ***Fetal fibronectin***

Of the two studies that used the qfFN test, one study (Bruijn<sup>62</sup>) appeared to conduct the test in accordance with the manufacturer's guidance. The other study (APOSTEL-1<sup>42</sup>) froze the samples as reported in *Studies evaluating more than one index test*. 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 3 months.

### ***Time to test results***

Based on the manufacturers' 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 qfFN. Some of the studies provided information within their methods relating to the maximum time that test personnel had to wait before reading the result from the test. This was always in accordance with the manufacturers' guidance.

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

### ***Test failure rates***

Only two studies (Goyal *et al.*<sup>52</sup> and Tanir *et al.*<sup>55</sup>), both of which evaluated Actim Partus, reported information about test failures: Goyal *et al.*<sup>52</sup> reported that there were no invalid tests and Tanir *et al.*<sup>55</sup> reported that there were two cases in which 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.<sup>52,55</sup> It was not clear whether these two women had a preterm delivery within 48 hours or within 7 days (i.e. whether they were assigned as true positive or false positive); therefore, it was not possible to conduct sensitivity analyses when these two cases were assigned as negative test results.

### ***Frozen samples***

Following the protocol amendment to include studies using frozen samples, two additional studies (APOSTEL-1<sup>42</sup> and Cooper *et al.*<sup>49</sup>) were included. Methodological details on freezing of the samples are described in the previous sections *Actim Partus* and *PartoSure*, and are further discussed in *Quality appraisal of included studies*. 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 the time frame suggested by the manufacturer's guidance, and is, therefore, unlikely to have a major impact on test accuracy. However, it should be noted that we have no published data to verify this.

## ***Description of included participants***

### **Studies evaluating more than one index test**

The APOSTEL-1 study, which evaluated both Actim Partus and qfFN in the same population, recruited 350 participants.<sup>42,43</sup> The characteristics of these participants are reported in *Table 4*. To summarise, the mean maternal age was 29.9 years [standard deviation (SD) 5.4 years] and the mean gestational age at presentation was 29.0 weeks (SD 2.7 weeks). In this study,<sup>42,43</sup> 20% of the participants ( $n = 71$ ) had multiple gestations and 23% ( $n = 79$ ) had previously delivered preterm.

The other study evaluating more than one index test in the same population was a much smaller study: Hadzi-Lega *et al.*<sup>44</sup> assessed both Actim Partus and PartoSure and recruited 57 participants. The characteristics of these participants are also reported in *Table 4*. In this study,<sup>44</sup> the median maternal age was reported as 27 years [interquartile range (IQR) 23.0–30.5 years] and the median gestational age at presentation was reported as 31 weeks (IQR 28.8–32.4 weeks).

### **Actim Partus**

In the 16 Actim Partus studies (including APOSTEL-1<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup>),<sup>42–44,57,58</sup> sample sizes ranged from 30 in Vishwekar *et al.*<sup>58</sup> to 468 in Tripathi *et al.*<sup>57</sup> Reported maternal ages ranged from a mean of 24.5 years (SD 5.16 years) in Altinkaya *et al.*<sup>46</sup> to a median of 31 years (IQR 28–34 years) in Danti *et al.*<sup>50</sup>

Gestation ranged from a mean of 28.7 weeks (SD not reported) in Riboni *et al.*<sup>54</sup> to 32.8 weeks (SD 3.24 weeks) in Goyal *et al.*<sup>52</sup> Further details describing the participant characteristics in each study are given in *Table 4*.

Three<sup>43,49,58</sup> of the 16 studies assessing Actim Partus included participants with multiple gestations (20%,  $n = 71$ ;<sup>43</sup> 6%,  $n = 20$ ;<sup>49</sup> and 7%,  $n = 258$ ). Previous preterm delivery was reported in seven studies,<sup>43,48,49,51–53,58</sup> and prevalence ranged from 3.9% ( $n = 2$ ) in Eroglu *et al.*<sup>51</sup> to 30% ( $n = 18$ ) in Goyal *et al.*<sup>52</sup>

### PartoSure

In the four PartoSure studies,<sup>41,44,59–61</sup> sample sizes ranged from 41 in Werlen *et al.*<sup>41</sup> to 203 in Nikolova *et al.*<sup>61</sup> Reported maternal age ranged from a median of 25 years (IQR 23–38 years) in Bolotskikh *et al.*<sup>59</sup> to a median of 27 years in Nikolova *et al.*<sup>60,61</sup> (range 18–43 years) and Hadzi-Lega *et al.*<sup>44</sup> (IQR 23–30.5 years). Gestational ages ranged from a mean of 29.5 weeks (SD 2.91 weeks) in Werlen *et al.*<sup>41</sup> to a median of 32 weeks (range 20.5–36.6 weeks) in Nikolova *et al.*<sup>60,61</sup> Further details describing the participant characteristics in each study are given in *Table 4*.

### Quantitative fetal fibronectin

In the two studies assessing qfFN (APOSTEL-1<sup>42</sup> and Bruijn<sup>62</sup>), sample sizes were 350 and 455, respectively. Maternal age was similar in these two studies [mean 29.9 years (SD 5.4 years) in APOSTEL-1<sup>42</sup> and mean 29.5 years (SD 5.2 years) in Bruijn<sup>62</sup>], as was gestation [mean 29.0 weeks (SD 2.7 weeks) in APOSTEL-1<sup>42</sup> and median 29.6 weeks (IQR 26.7–31.6 weeks) in Bruijn<sup>62</sup>]. The APOSTEL-1 study<sup>42</sup> had a higher proportion of multiple pregnancies than Bruijn<sup>62</sup> [20% ( $n = 71$ ) in APOSTEL-1<sup>42</sup> vs. 15% ( $n = 67$ ) in Bruijn<sup>62</sup>]. Proportionally more women had previously delivered preterm in APOSTEL-1<sup>42</sup> than in Bruijn<sup>62</sup> [23% ( $n = 79$ ) in APOSTEL-1<sup>42</sup> vs. 16% ( $n = 72$ ) in Bruijn<sup>62</sup>]. Further details describing the participant characteristics in each study are given in *Table 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 48 hours and/or 7 days). For this reason, it is important to consider the use of any treatments that might have an impact on the reference standard and, thus, the test accuracy results. Whether or not a woman received treatment for symptoms of preterm labour varied substantially between studies. This variability was based on (1) the standard treatment protocols used within each study and (2) the test(s) used to initiate treatment within the study (e.g. Actim Partus, PartoSure or qfFN or another test that is not being assessed as part of this review, such as transvaginal cervical length). Subsequently, when 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 had an impact on the test accuracy results.

### Studies evaluating more than one index test

Treatment decisions from the APOSTEL-1 trial<sup>42,43</sup> were not based on the test results of Actim Partus or qfFN alone. Instead, treatment decisions were based on the combined results of two tests not being assessed in this review (cervical length measurement and qfFN with a threshold of 50 ng/ml). In addition, 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 of < 10 mm but not for women with a cervical length of > 30 mm. Women with a cervical length between 10 mm and 30 mm and a positive fFN result were encouraged to receive tocolytics, whereas those with a negative fFN result 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.

Treatment decisions from the study by Hadzi-Lega *et al.*<sup>44</sup> (a single-centre study assessing both Actim Partus and PartoSure) followed the standard of care at the hospital. This included hospitalisation, 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. Thirty-eight of 57 women (67%) received tocolytics (note that 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.<sup>44</sup>

### Actim Partus

Sixteen studies assessed Actim Partus, two of which (APOSTEL-1<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup>) also assessed another index test and have been discussed previously (see *Studies evaluating more than one index test*). The remaining 14 studies<sup>45–58</sup> all report the use of tocolytics and most specify the type (typically Atosiban, calcium channel blockers, nifedipine, ritodrine or magnesium sulphate). All of these studies except Azlin *et al.*<sup>47</sup> report the use of corticosteroids (either betamethasone or dexamethasone, when specified).

One study reports that cervical length measurements were available to managing clinicians and were used to aid treatment decisions.<sup>50</sup> Of the 60 patients in this study, 22 (37%) received tocolytics and 28 (47%) received corticosteroids.

In another study,<sup>51</sup> it is reported that patients were admitted to hospital based on frequency of contractions and digital examination. In this study, women at > 34 weeks' gestation 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 who received corticosteroids.<sup>51</sup> In a further four studies,<sup>53,54,57,58</sup> treatment decisions were also guided by gestational age. In Lembet *et al.*<sup>53</sup> patients at > 34 weeks' gestation were not administered tocolytics, those at > 24 weeks' gestation received corticosteroids and those at < 28 weeks' gestation 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.<sup>53</sup> In Riboni *et al.*,<sup>54</sup> women at < 34 weeks' gestation received corticosteroids, and all patients received tocolytics. In the studies by Tripathi *et al.*<sup>57</sup> and Vishwekar *et al.*,<sup>58</sup> women at < 34 weeks' gestation received tocolytics and corticosteroids. The studies by Riboni *et al.*,<sup>54</sup> Tripathi *et al.*<sup>57</sup> and Vishwekar *et al.*<sup>58</sup> do not report the proportion of patients who received these treatments (or the Actim Partus results of these patients).

Five studies state that treatment decisions were based on standard hospital protocol.<sup>47–49,52,56</sup> In these studies, it is largely unclear whether diagnostic tests, symptoms or maternal characteristics (e.g. gestation) were used to guide decision-making. Four of these studies report details on the proportion of patients receiving treatment. In the study by Azlin *et al.*,<sup>47</sup> 12 out of 51 patients (24%) received tocolytics (two of whom had a positive Actim Partus result), but no details were reported regarding how many patients received corticosteroids. In the study by Brik *et al.*,<sup>48</sup> 213 out of 276 patients (77%) received tocolytics and 200 (73%) received corticosteroids. In the study by Cooper *et al.*,<sup>49</sup> 8 out of 349 patients received tocolytics (2%) and 56 (16%) received corticosteroids. In the study by Goyal *et al.*,<sup>52</sup> all patients received tocolytics and the number receiving corticosteroids was unclear. The number of patients receiving tocolytics or corticosteroids was not reported in Ting *et al.*<sup>56</sup>

The remaining three studies report limited detail regarding how treatment decisions were made.<sup>45,46,55</sup> In the studies by Abo El-Ezz *et al.*<sup>45</sup> and Altinkaya *et al.*,<sup>46</sup> all patients were admitted to hospital. Abo El-Ezz *et al.*<sup>45</sup> report that all patients received tocolytics, and it is unclear how many patients received corticosteroids. In the study by Altinkaya *et al.*,<sup>46</sup> patients whose symptoms persisted received tocolytics, and all patients were administered corticosteroids. Neither of these studies report details regarding the Actim Partus results of those receiving treatment. In the study by Tanir *et al.*,<sup>55</sup> 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.<sup>55</sup>

### PartoSure

Four studies assessed PartoSure, one of which<sup>44</sup> also assessed another index test and has been discussed previously (see *Studies evaluating more than one index test*).

In one of the remaining three studies,<sup>59</sup> tocolytics were given irrespective of test outcomes (all admitted women were treated). In another study,<sup>41</sup> treatment was at the discretion of the investigator in accordance with the protocol of the department. Of the 41 women in this study, 13 received corticosteroids and 25 received tocolytics, but no details were provided for treatment administration in connection with the PartoSure test results. In the remaining study,<sup>60,61</sup> no treatment protocols were reported.

### Quantitative fetal fibronectin

Two studies assessed qfFN, one of which (APOSTEL-1<sup>42,43</sup>) also assessed another index test and has been discussed previously (see *Studies evaluating more than one index test*).<sup>42–44</sup>

In the other study (Bruijn<sup>62</sup>), treatment with tocolytics and steroids was based on a combination of cervical length and the qfFN result at a threshold of 50 ng/ml, neither of which are index tests in this review. In addition, no data were provided on how many women received treatments.

### Quality appraisal of included studies

Quality appraisal was conducted, using phase 3 of the QUADAS-2 tool,<sup>39</sup> 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<sup>42,43</sup> and the study by Hadzi-Lega *et al.*<sup>44</sup> each assess two index tests (Actim Partus and qfFN, and PartoSure and Actim Partus, respectively). For these two studies, the index test domain was conducted separately for each test.

It is important to note that the QUADAS-2<sup>39</sup> tool assesses the likely risk of bias and not the presence or magnitude of bias. Therefore, any rating of 'high' 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 following sections, and a summary of the QUADAS-2 ratings is provided in *Table 5*. Further details relating to patient selection, index tests, reference standard, and flow and timing can also be found in *Table 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<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup>) evaluated more than one index test in the same population. 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. owing to differences in the recruited populations, hospitals and other factors that may have an impact on whether or not a woman is likely to have a preterm 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 regarding whether or not the selection of women in these studies could have introduced bias. Indeed, only five studies, four of which were Actim Partus studies<sup>47,48,50,58</sup> and one of which was a PartoSure study,<sup>60,61</sup> clearly reported selection procedures.

In almost all of the studies, index tests were conducted in a manner that is consistent with clinical practice. However, both the APOSTEL-1 study<sup>42,43</sup> and the study by Cooper *et al.*<sup>49</sup> used frozen samples, which means that the index tests (Actim Partus in both studies and qfFN in APOSTEL-1) were interpreted after the reference standard. Owing to the nature of the tests, this is unlikely to have much impact, although some amount of bias cannot be completely ruled out. In addition, in APOSTEL-1<sup>42,43</sup> and the study by Ting *et al.*,<sup>56</sup> 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.

Furthermore, although not likely to have a major impact on 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





Quality appraisal questions	Test																			
	Actim Partus and fFN <sup>a</sup>	Actim Partus and PartoSure <sup>b</sup>	Actim Partus														PartoSure	fFN		
	APOSTEL-1 <sup>42,43</sup>	Hadzi-Lega <i>et al.</i> <sup>44</sup> (2017)	Abo El-Ezz <i>et al.</i> <sup>45</sup> (2014)	Altinkaya <i>et al.</i> <sup>46</sup> (2009)	Azlin <i>et al.</i> <sup>47</sup> (2010)	Brik <i>et al.</i> <sup>48</sup> (2010)	Cooper <i>et al.</i> <sup>49</sup> (2012)	Danti <i>et al.</i> <sup>50</sup> (2011)	Eroglu <i>et al.</i> <sup>51</sup> (2007)	Goyal <i>et al.</i> <sup>52</sup> (2016)	Lembet <i>et al.</i> <sup>53</sup> (2002)	Riboni <i>et al.</i> <sup>54</sup> (2011)	Tanir <i>et al.</i> <sup>55</sup> (2009)	Ting <i>et al.</i> <sup>56</sup> (2007)	Tripathi <i>et al.</i> <sup>57</sup> (2016)	Vishwekar <i>et al.</i> <sup>58</sup> (2017)	Bolotskikh <i>et al.</i> <sup>59</sup> (2017)	Nikolova <i>et al.</i> <sup>60,61</sup> (2014 and 2015)	Werlen <i>et al.</i> <sup>41</sup> (2015)	Bruijn <sup>62</sup>
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 <sup>c</sup>	L	L	L	L	L	L	L	L	L	L	L	L	H <sup>c</sup>	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 prespecified?	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

continued



Quality appraisal questions	Test																			
	Actim Partus and fFN <sup>a</sup>	Actim Partus and PartoSure <sup>b</sup>	Actim Partus														PartoSure	fFN		
	APOSTEL-1 <sup>42,43</sup>	Hadzi-Lega <i>et al.</i> <sup>44</sup> (2017)	Abo El-Ezz <i>et al.</i> <sup>45</sup> (2014)	Altinkaya <i>et al.</i> <sup>46</sup> (2009)	Azlin <i>et al.</i> <sup>47</sup> (2010)	Brik <i>et al.</i> <sup>48</sup> (2010)	Cooper <i>et al.</i> <sup>49</sup> (2012)	Danti <i>et al.</i> <sup>50</sup> (2011)	Eroglu <i>et al.</i> <sup>51</sup> (2007)	Goyal <i>et al.</i> <sup>52</sup> (2016)	Lembet <i>et al.</i> <sup>53</sup> (2002)	Riboni <i>et al.</i> <sup>54</sup> (2011)	Tanir <i>et al.</i> <sup>55</sup> (2009)	Ting <i>et al.</i> <sup>56</sup> (2007)	Tripathi <i>et al.</i> <sup>57</sup> (2016)	Vishwekar <i>et al.</i> <sup>58</sup> (2017)	Bolotskikh <i>et al.</i> <sup>59</sup> (2017)	Nikolova <i>et al.</i> <sup>60,61</sup> (2014 and 2015)	Werlen <i>et al.</i> <sup>41</sup> (2015)	Bruijn <sup>62</sup>
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
H, high; L, low; N, no; N/A, not applicable; U, unknown; Y, yes.																				
a Index test 1 = Actim Partus, index test 2 = fFN.																				
b Index test 1 = Actim Partus, index test 2 = PartoSure.																				
c The samples were taken from the posterior fornix of the vagina rather than the external cervical os.																				

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.<sup>42–44,49–51,53,55,56,60–62</sup>

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<sup>42,43</sup> and Cooper *et al.*<sup>49</sup>). 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<sup>41–45,48,52,56,57,60–62</sup> were rated as being of high risk of bias regarding not including all women in analyses. A 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 in which a high risk of bias rating has been given (see *Table 5*).

### Results of quantitative data synthesis (test accuracy data)

#### The 7-day delivery reference standard

##### Studies evaluating more than one index test

Two studies (APOSTEL-1<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup>) reported test accuracy data on two index tests. Both studies used only the 7-day (and not the 48-hour) delivery reference standard. The prevalence of preterm birth within 7 days was 19.7% (95% CI 15.7% to 24.3%) in the APOSTEL-1 study<sup>42,43</sup> and 10.5% (95% CI 4.0% to 21.5%) in the study by Hadzi-Lega *et al.*<sup>44</sup>

The APOSTEL-1 study reports test accuracy from 350 women for both Actim Partus and qfFN (*Table 6*).<sup>42,43</sup> The sensitivity and specificity for Actim Partus, against delivery within 7 days, were 78.3% (95% CI 66.7% to 87.3%) and 89.3% (95% CI 85.1% to 92.7%), respectively. As would be expected, in the qfFN 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 qfFN 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, at which sensitivity was 71.0% (95% CI 58.8% to 83.1%) and specificity was 83.6% (95% CI 78.8% to 87.8%). With regard to the qfFN data from APOSTEL-1, the threshold with the highest PPV was 500 ng/ml (70.7%, 95% CI 54.5% to 83.9%) and the threshold with the lowest PPV was 10 ng/ml (28.9%, 95% CI 23.2% to 35.3%). For Actim Partus the PPV of delivery within 7 days was 64.3% (95% CI 53.1% to 74.4%). For qfFN, the highest NPV was at a threshold of 10 ng/ml (97.5%, 95% CI 93.0% to 99.5%) and the lowest at a threshold of 500 ng/ml (87.1%, 95% CI 82.8% to 90.6%). For Actim Partus, NPV was 94.4% (95% CI 90.9% to 96.8%). LRs, 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 the study by Hadzi-Lega *et al.*,<sup>44</sup> in which test accuracy data from 57 women were reported for both Actim Partus and PartoSure (see *Table 6*). The sensitivity of both Actim Partus and PartoSure for delivery within 7 days was 83.3% (95% CI 35.9% to 99.6%), and specificity was higher for PartoSure (90.2%, 95% CI 78.6% to 96.7%) than for Actim Partus (76.5%, 95% CI 62.5% to 87.2%). In addition, PPV, LR+ and concordance were higher for PartoSure than for Actim Partus, although the wide CIs, particularly for PPV, are notable. LR– and NPV were similar for both tests and diagnostic yield was higher for Actim Partus than for PartoSure. Specific values are given in *Table 6*.

##### Actim Partus

Results for Actim Partus against the 7-day delivery reference standard were provided by 16 studies (*Figure 3*, and see *Table 6*). Across these studies, sensitivity ranged from 33.3% (95% CI 4.3% to 77.7%) in the study by Cooper *et al.*<sup>49</sup> to 94.7% (95% CI 89.9% to 97.7%) in the study by Tripathi *et al.*<sup>57</sup> Specificity of Actim Partus ranged from 50.0% (95% CI 24.7% to 75.3%) in the study by Goyal *et al.*<sup>52</sup> to 93.5% (95% CI

**TABLE 6** Calculated diagnostic accuracy parameters against the 7-day reference standard

Study (first author/study name and year)	Participants (n)	Diagnostic accuracy (95% CI)								
		Sensitivity (%)	Specificity (%)	LR+	LR-	PPV (%)	NPV (%)	Prevalence (%)	Concordance	Yield
<b>Study assessing qfFN and Actim Partus</b>										
APOSTEL-1 (2016) <sup>42,43</sup>										
ffN at 10 ng/ml	350	95.7 (87.8 to 99.1)	42.3 (36.5 to 48.4)	1.66 (1.48 to 1.86)	0.10 (0.03 to 0.31)	28.9 (23.2 to 35.3)	97.5 (93.0 to 99.5)	19.7 (15.7 to 24.3)	0.53 (0.48 to 0.58)	0.65 (0.60 to 0.70)
ffN at 200 ng/ml	350	71.0 (58.8 to 81.3)	83.6 (78.8 to 87.8)	4.34 (3.20 to 5.88)	0.35 (0.24 to 0.50)	51.6 (41.1 to 62.0)	92.2 (88.1 to 95.1)	19.7 (15.7 to 24.3)	0.81 (0.77 to 0.85)	0.27 (0.23 to 0.32)
ffN at 500 ng/ml	350	42.0 (30.2 to 54.5)	95.7 (92.7 to 97.8)	9.84 (5.30 to 18.28)	0.61 (0.49 to 0.74)	70.7 (54.5 to 83.9)	87.1 (82.8 to 90.6)	19.7 (15.7 to 24.3)	0.85 (0.81 to 0.89)	0.12 (0.09 to 0.16)
Actim Partus	350	78.3 (66.7 to 87.3)	89.3 (85.1 to 92.7)	7.33 (5.11 to 10.51)	0.24 (0.16 to 0.38)	64.3 (53.1 to 74.4)	94.4 (90.9 to 96.8)	19.7 (15.7 to 24.3)	0.87 (0.83 to 0.91)	0.24 (0.20 to 0.29)
<b>Study assessing PartoSure and Actim Partus</b>										
Hadzi-Lega (2017) <sup>44</sup>										
PartoSure	57	83.3 (35.9 to 99.6)	90.2 (78.6 to 96.7)	8.50 (3.43 to 21.03)	0.18 (0.03 to 1.11)	50.0 (18.7 to 81.3)	97.9 (88.7 to 99.9)	10.5 (4.0 to 21.5)	0.90 (0.79 to 0.96)	0.18 (0.09 to 0.30)
Actim Partus	57	83.3 (35.9 to 99.6)	76.5 (62.5 to 87.2)	3.54 (1.92 to 6.52)	0.22 (0.04 to 1.31)	29.3 (10.3 to 56.0)	97.5 (86.8 to 99.9)	10.5 (4.0 to 21.5)	0.77 (0.64 to 0.87)	0.30 (0.18 to 0.43)
<b>Actim Partus</b>										
Abo El-Ezz (2014) <sup>45</sup>										
Actim Partus	57	66.7 (47.2 to 82.7)	66.7 (46.0 to 83.5)	2.00 (1.11 to 3.61)	0.50 (0.28 to 0.89)	69.0 (49.2 to 84.7)	64.3 (44.1 to 81.4)	52.6 (39.0 to 66.0)	0.67 (0.53 to 0.79)	0.51 (0.37 to 0.64)
Altinkaya (2009) <sup>46</sup>										
Actim Partus	105	64.3 (35.1 to 87.2)	82.4 (73.0 to 89.6)	3.66 (2.02 to 6.61)	0.43 (0.21 to 0.88)	36.0 (18.0 to 57.5)	93.8 (86.0 to 97.9)	13.3 (7.5 to 21.4)	0.80 (0.71 to 0.87)	0.24 (0.16 to 0.33)
Azlin (2010) <sup>47</sup>										
Actim Partus	51	80.0 (28.4 to 99.5)	93.5 (82.1 to 98.6)	12.27 (3.77 to 39.86)	0.21 (0.04 to 1.24)	57.1 (18.4 to 90.1)	97.7 (88.0 to 99.9)	9.8 (3.3 to 21.4)	0.92 (0.81 to 0.98)	0.14 (0.06 to 0.26)
Brik (2010) <sup>48</sup>										
Actim Partus	276	74.2 (55.4 to 88.1)	66.1 (59.8 to 72.0)	2.19 (1.67 to 2.87)	0.39 (0.21 to 0.71)	21.7 (14.3 to 30.8)	95.3 (90.9 to 97.9)	11.2 (7.8 to 15.6)	0.67 (0.61 to 0.73)	0.38 (0.33 to 0.44)

continued

**TABLE 6** Calculated diagnostic accuracy parameters against the 7-day reference standard (*continued*)

Study (first author/study name and year)	Participants (n)	Diagnostic accuracy (95% CI)								
		Sensitivity (%)	Specificity (%)	LR+	LR-	PPV (%)	NPV (%)	Prevalence (%)	Concordance	Yield
Cooper (2012) <sup>49</sup>										
Actim Partus	349	33.3 (4.3 to 77.7)	74.1 (69.1 to 78.6)	1.28 (0.41 to 4.04)	0.90 (0.51 to 1.59)	2.2 (0.3 to 7.7)	98.4 (96.1 to 99.6)	1.7 (0.6 to 3.7)	0.73 (0.68 to 0.78)	0.26 (0.22 to 0.31)
Danti (2011) <sup>50</sup>										
Actim Partus	60	50.0 (6.8 to 93.2)	69.6 (55.9 to 81.2)	1.65 (0.57 to 4.74)	0.72 (0.27 to 1.94)	10.5 (1.3 to 33.1)	95.1 (83.5 to 99.4)	6.7 (1.8 to 16.2)	0.68 (0.55 to 0.80)	0.32 (0.20 to 0.45)
Eroglu (2007) <sup>51</sup>										
Actim Partus	51	83.3 (35.9 to 99.6)	84.4 (70.5 to 93.5)	5.36 (2.48 to 11.56)	0.20 (0.03 to 1.19)	41.7 (15.2 to 72.3)	97.4 (86.5 to 99.9)	11.8 (4.4 to 23.9)	0.84 (0.71 to 0.93)	0.24 (0.13 to 0.38)
Goyal (2016) <sup>52</sup>										
Actim Partus	60	59.1 (43.2 to 73.7)	50.0 (24.7 to 75.3)	1.18 (0.68 to 2.04)	0.82 (0.45 to 1.50)	76.5 (58.8 to 89.3)	30.8 (14.3 to 51.8)	73.3 (60.3 to 83.9)	0.57 (0.43 to 0.69)	0.57 (0.43 to 0.69)
Lembet (2002) <sup>53</sup>										
Actim Partus	36	93.8 (69.8 to 99.8)	85.0 (62.1 to 96.8)	6.25 (2.19 to 17.88)	0.07 (0.01 to 0.49)	83.3 (58.6 to 96.4)	94.4 (72.7 to 99.9)	44.4 (27.9 to 61.9)	0.89 (0.74 to 0.97)	0.50 (0.33 to 0.67)
Riboni (2011) <sup>54</sup>										
Actim Partus	210	50.0 (15.7 to 84.3)	83.7 (77.8 to 88.5)	3.06 (1.43 to 6.54)	0.60 (0.30 to 1.20)	10.8 (3.0 to 25.4)	97.7 (94.2 to 99.4)	3.8 (1.7 to 7.4)	0.82 (0.77 to 0.87)	0.18 (0.13 to 0.24)
Tanir (2009) <sup>55</sup>										
Actim Partus	68	93.3 (68.1 to 99.8)	79.2 (65.9 to 89.2)	4.50 (2.61 to 7.74)	0.08 (0.01 to 0.56)	56.0 (34.9 to 75.6)	97.7 (87.7 to 99.9)	22.1 (12.9 to 33.8)	0.82 (0.71 to 0.91)	0.37 (0.25 to 0.49)
Ting (2007) <sup>56</sup>										
Actim Partus	94	70.6 (44.0 to 89.7)	77.9 (67.0 to 87.6)	3.20 (1.90 to 5.38)	0.38 (0.18 to 0.80)	41.4 (23.5 to 61.1)	92.3 (83.0 to 97.5)	18.1 (10.9 to 27.4)	0.77 (0.67 to 0.85)	0.31 (0.22 to 0.41)
Tripathi (2016) <sup>57</sup>										
Actim Partus	468	94.7 (89.9 to 97.7)	92.4 (88.9 to 95.1)	12.43 (8.45 to 18.30)	0.06 (0.03 to 0.11)	85.7 (79.5 to 90.6)	97.3 (94.8 to 98.8)	32.5 (28.3 to 37.0)	0.93 (0.91 to 0.95)	0.36 (0.32 to 0.41)
Vishwekar (2017) <sup>58</sup>										
Actim Partus	30	72.2 (46.5 to 90.3)	90.9 (58.7 to 99.8)	7.94 (1.20 to 52.62)	0.31 (0.14 to 0.66)	92.9 (66.1 to 99.8)	66.7 (38.4 to 88.2)	62.1 (42.3 to 79.3)	0.79 (0.60 to 0.92)	0.48 (0.29 to 0.67)

Study (first author/study name and year)	Participants (n)	Diagnostic accuracy (95% CI)								
		Sensitivity (%)	Specificity (%)	LR+	LR-	PPV (%)	NPV (%)	Prevalence (%)	Concordance	Yield
<b>PartoSure</b>										
Bolotskikh (2017) <sup>59</sup>										
PartoSure	99	100.0 (73.5 to 100.0)	95.4 (88.6 to 98.7)	21.75 (8.35 to 56.64)	0.00 (N/A)	75.0 (47.6 to 92.7)	100.0 (95.7 to 100.0)	12.1 (6.4 to 20.2)	0.96 (0.90 to 0.99)	0.16 (0.10 to 0.25)
Nikolova (2014 and 2015) <sup>60,61</sup>										
PartoSure	203	80.0 (63.1 to 91.6)	94.6 (90.1 to 97.5)	14.93 (7.74 to 28.80)	0.21 (0.11 to 0.41)	75.7 (58.8 to 88.2)	95.8 (91.5 to 98.3)	17.2 (12.3 to 23.2)	0.92 (0.88 to 0.95)	0.18 (0.13 to 0.24)
Werlen (2015) <sup>41</sup>										
PartoSure	41	0.0 (0.0 to 97.5)	97.5 (96.8 to 99.9)	0.00 (N/A)	1.03 (0.98 to 1.08)	0.0 (0.0 to 97.5)	97.5 (96.8 to 99.9)	2.4 (0.1 to 12.9)	0.95 (0.84 to 0.99)	0.02 (0.00 to 0.13)
<b>qfFN</b>										
Bruijn (2016) <sup>62</sup>										
fFN at 10 ng/ml	455	93.8 (82.8 to 98.7)	32.2 (27.7 to 37.0)	1.38 (1.25 to 1.53)	0.19 (0.06 to 0.59)	14.0 (10.4 to 18.3)	97.8 (93.6 to 99.5)	10.5 (7.9 to 13.7)	0.39 (0.34 to 0.43)	0.71 (0.66 to 0.75)
fFN at 200 ng/ml	455	70.8 (55.9 to 83.0)	78.6 (74.3 to 82.5)	3.31 (2.55 to 4.30)	0.37 (0.24 to 0.58)	28.1 (20.3 to 37.0)	95.8 (93.1 to 97.7)	10.5 (7.9 to 13.7)	0.78 (0.74 to 0.82)	0.27 (0.23 to 0.31)
fFN at 500 ng/ml	455	29.2 (17.0 to 44.1)	94.3 (91.6 to 96.4)	5.16 (2.85 to 9.34)	0.75 (0.63 to 0.90)	37.8 (22.5 to 55.2)	91.9 (88.8 to 94.3)	10.5 (7.9 to 13.7)	0.99 (0.84 to 0.90)	0.08 (0.06 to 0.11)
N/A, not applicable.										

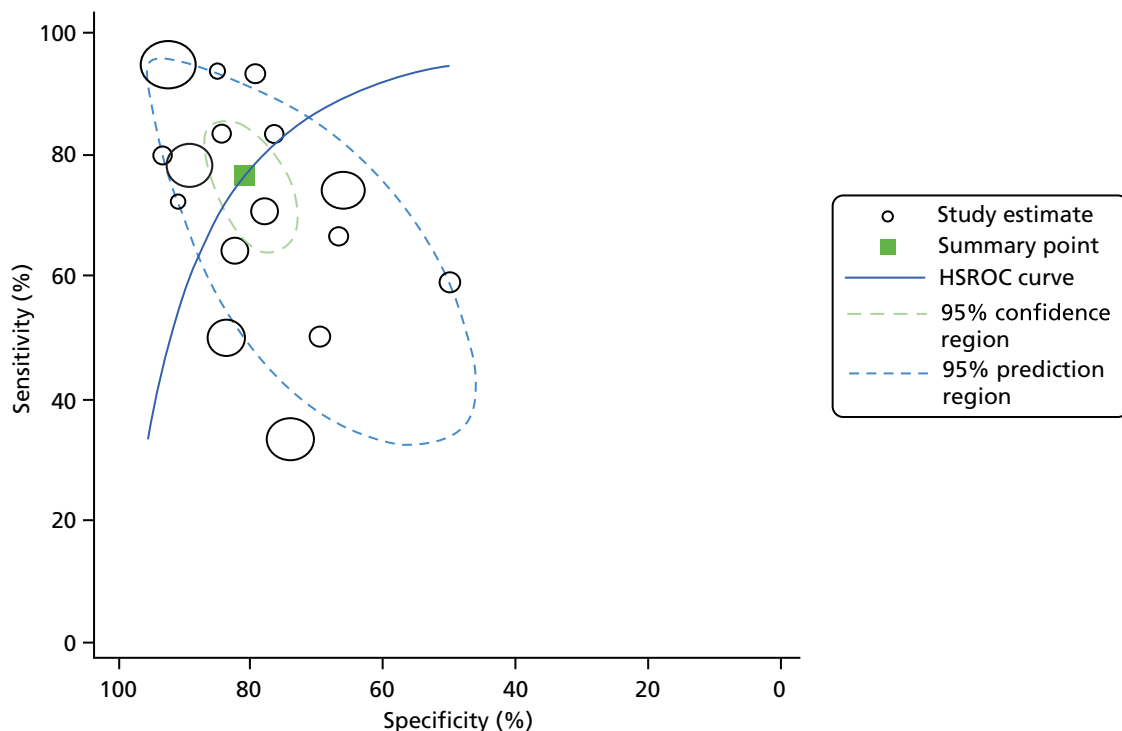
82.1% to 98.6%) in the study by Azlin *et al.*<sup>47</sup> The three studies with the lowest sensitivity were those by Cooper *et al.*<sup>49</sup> (33.3%, 95% CI 4.3% to 77.7%), Danti *et al.*<sup>50</sup> (50%, 95% CI 6.8% to 93.2%) and Riboni *et al.*<sup>54</sup> (50%, 95% CI 15.7% to 84.3%). The prevalence reported in these studies was much lower [prevalence ranging from 1.7% (95% CI 0.6% to 3.7%) to 6.7% (95% CI 1.8% to 16.2%)] than in all other studies [prevalence ranging from 9.8% (95% CI 3.3% to 21.4%) to 73.3% (95% CI 60.3% to 83.9%)]. Indeed, the large range of prevalence estimates across these studies is particularly noteworthy (see *Differences between studies in prevalence of preterm birth*). Meanwhile, the three studies with the lowest specificities were those by Goyal *et al.*<sup>54</sup> (50%, 95% CI 24.7% to 75.3%), Brik *et al.*<sup>48</sup> (66%, 95% CI 59.8% to 72.0%) and Abo El-Ezz *et al.*<sup>45</sup> (67%, 95% CI 46.0% to 83.5%). There were no obvious methodological or participant characteristics in these studies to explain the differences, and, although two of these studies<sup>45,54</sup> had high prevalence, the other<sup>48</sup> did not.

A summary ROC plot for all 16 studies assessing Actim Partus against the 7-day delivery reference standard is provided in *Figure 3*. Pooled analyses were undertaken for these data and provided a pooled sensitivity of 77% (95% CI 68% to 83%) and a pooled specificity of 81% (95% CI 76% to 85%).

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*.

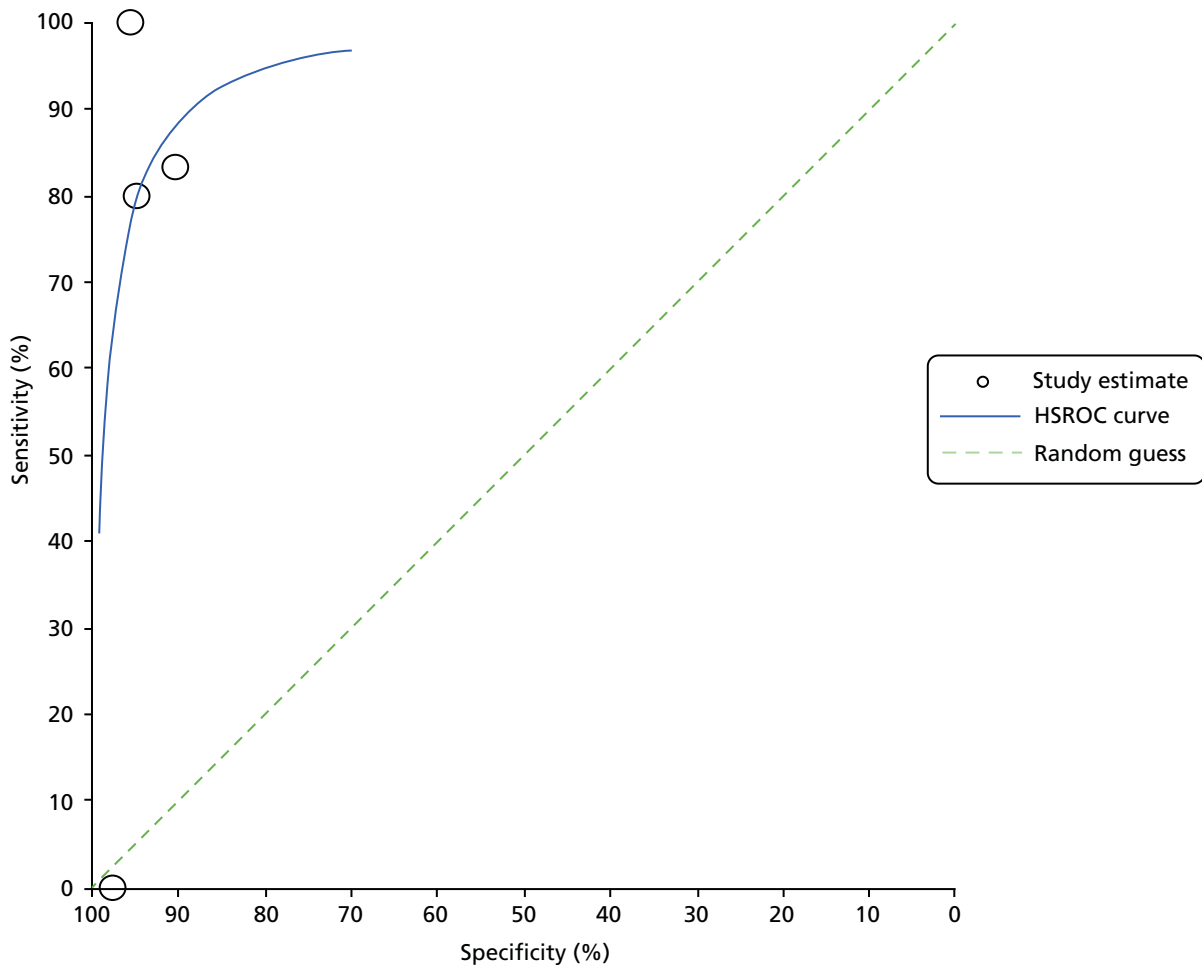
### PartoSure

Results for PartoSure against the 7-day delivery reference standard were reported in four studies (*Figure 4*, and see *Table 6*). Prevalence of preterm delivery within 7 days ranged from 2.4% (95% CI 0.1% to 12.9%) in the study by Werlen *et al.*<sup>41</sup> to 17.2% (95% CI 12.3% to 23.2%) in the study by Nikolova *et al.*<sup>61</sup> These four studies had wide-ranging sensitivity, from 0% (95% CI 0.0% to 97.5%) in the study by Werlen *et al.*<sup>41</sup> to 100% (95% CI 73.5% to 100.0%) in the study by Bolotskikh *et al.*<sup>59</sup> whereas specificity was more similar across studies, ranging from 90.2% (95% CI 78.6% to 96.7%) in the study by Hadzi-Lega *et al.*<sup>44</sup> to 97.5% (95% CI 96.8% to 99.9%) in the study by Werlen *et al.*<sup>41</sup> The low sensitivity in the study by Werlen *et al.*<sup>41</sup> is because, in the sample of 41 participants, only one tested (falsely) positive using the PartoSure test.



**FIGURE 3** The ROC plot for Actim Partus against the 7-day reference standard. HSROC, hierarchical summary receiver-operating characteristic.





**FIGURE 4** The ROC plot for PartoSure against the 7-day reference standard.

Discounting this study, the sensitivity range would be 80% (95% CI 63.1% to 91.6%) in the study by Nikolova *et al.*<sup>61</sup> to 100% (95% CI 73.5% to 100.0%) in the study by Bolotskikh *et al.*<sup>59</sup>

A summary ROC plot for the four studies assessing PartoSure against the 7-day delivery reference standard is provided in *Figure 3*. Pooled analyses were conducted for these data and provided a pooled sensitivity of 83% (95% CI 61% to 94%) and a pooled specificity of 95% (95% CI 89% to 98%).

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*.

### Quantitative fetal fibronectin

Results against the 7-day delivery reference standard for the two qfFN studies (Bruijn<sup>62</sup> and APOSTEL-1<sup>42,43</sup>), at the three thresholds (10 ng/ml, 200 ng/ml and 500 ng/ml), are presented in *Table 6*. Prevalence of preterm birth within the sample was lower in Bruijn<sup>62</sup> (10.5%, 95% CI 7.9% to 13.7%) than in APOSTEL-1<sup>42,43</sup> (19.7%, 95% CI 15.7% to 24.3%). Bruijn<sup>62</sup> presented with slightly lower (within 2%) sensitivity values than APOSTEL-1<sup>42,43</sup> at both the 10-ng/ml and 200-ng/ml thresholds (see *Table 6*). At the 500-ng/ml threshold, the sensitivity was much lower in Bruijn<sup>62</sup> (29.2%, 95% CI 17.0% to 44.1%) than in APOSTEL-1<sup>42,43</sup> (42%, 95% CI 30.2% to 54.4%). Similarly, specificity values were slightly lower (within 5%) in Bruijn<sup>62</sup> than in APOSTEL-1<sup>42,43</sup> at both the 200-ng/ml and 500-ng/ml thresholds, whereas at the 10-ng/ml threshold the specificity was much lower (32.2%, 95% CI 27.7% to 37.0%) in Bruijn<sup>62</sup> than in APOSTEL-1<sup>42,43</sup> (42.3%, 95% CI 36.5% to 48.4%).

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

### The 48-hour delivery reference standard

Seven of the included studies also provided test accuracy data for the index tests against a 48-hour preterm delivery reference standard. Six of these studies evaluated Actim Partus<sup>48,52,53,56-58</sup> and one evaluated PartoSure.<sup>41</sup>

#### *Actim Partus*

Across the six studies evaluating Actim Partus against the prevalence of preterm birth within 48 hours,<sup>48,52,53,56-58</sup> the prevalence of preterm birth within 48 hours ranged from 5.3% (95% CI 1.7% to 12.0%) in the study by Ting *et al.*<sup>56</sup> to 58.3% (95% CI 44.9% to 70.9%) in the study by Goyal *et al.*<sup>52</sup> Sensitivity ranged from 65.7% (95% CI 47.8% to 80.9%) in the study by Goyal *et al.*<sup>52</sup> to 100.0% (95% CI 47.8% to 100.0%) in the study by Ting *et al.*<sup>56</sup> Specificity ranged from 56.0% (95% CI 34.9% to 75.6%) in the study by Goyal *et al.*<sup>52</sup> to 82.4% (95% CI 56.6% to 96.2%) in the study by Vishwekar *et al.*<sup>58</sup>

Specific sensitivity and specificity values for all six studies are given in *Table 7*, in which it can be seen that the sensitivity and specificity of Actim Partus for the 48-hour reference standard were lowest in the studies by Goyal *et al.*<sup>52</sup> and Brik *et al.*<sup>48</sup> compared with the other four studies,<sup>53,56-58</sup> which seem more in line with each other. There were no obvious methodological or participant characteristic differences in these studies (other than that the women in the studies by Goyal *et al.*<sup>52</sup> and Brik *et al.*<sup>48</sup> were, on average, 1 year older than those in the other studies). Brik *et al.*<sup>48</sup> 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 5*. Pooled analyses were conducted for these data and provided a pooled sensitivity of 87% (95% CI 74% to 94%) and a pooled specificity of 73% (95% CI 62% to 82%).

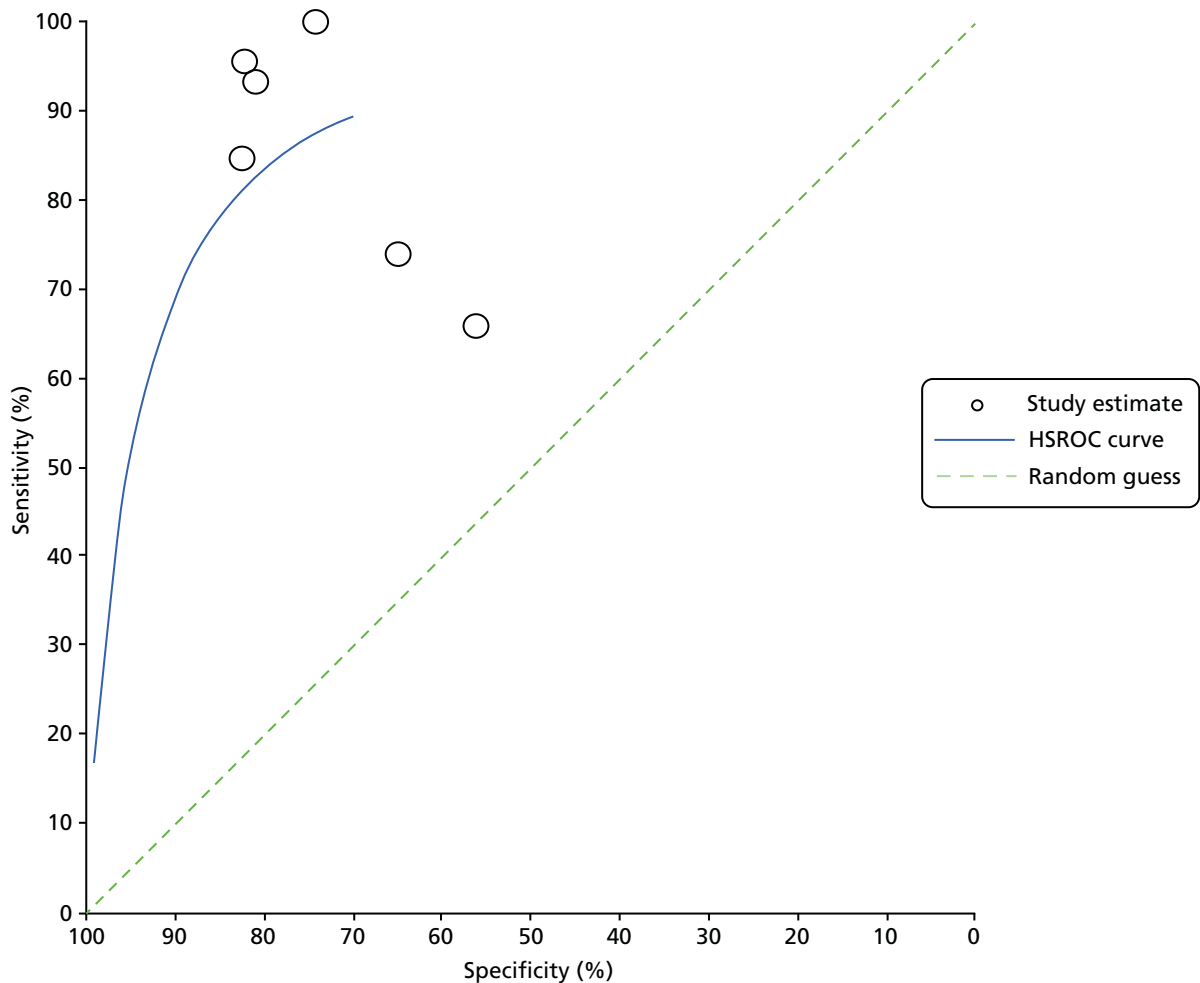
In the six studies evaluating Actim Partus against the prevalence of preterm birth within 48 hours, the PPV was lower in the studies by Brik *et al.*<sup>48</sup> and Ting *et al.*<sup>56</sup> [16.0% (95% CI 9.6% to 24.4%) and 17.9% (95% CI 6.1% to 36.9%), respectively] than in the other four studies [range 62.1% (95% CI 54.4% to 69.5%) to 78.6% (95% CI 49.2% to 95.3%)]. This is likely to be linked to the prevalence also being low in these two studies (8.3%, 95% CI 5.4% to 12.2% in the study by Brik *et al.*<sup>48</sup> and 5.3%, 95% CI 1.7% to 12.0% in the study by Ting *et al.*<sup>56</sup>). Conversely, NPV was lowest in the study by Goyal *et al.*<sup>52</sup> (53.8%, 95% CI 33.4% to 73.4%), with NPV in the other five studies ranging from 87.5% (95% CI 61.7% to 98.4%) in the study by Vishwekar *et al.*<sup>58</sup> to 100% (95% CI 94.6% to 100.0%) in the study by Ting *et al.*<sup>56</sup> Looking at the diagnostic yield, the study by Goyal *et al.*<sup>52</sup> was the only study in which > 50% of the population had a positive Actim Partus result (57%, 95% CI 43% to 69%); all other studies had a diagnostic yield of ≤ 50%. Data from these six Actim Partus studies were also used to calculate LR+, LR- and concordance. These values are provided in *Table 7*.

#### *PartoSure*

In the study evaluating PartoSure against the prevalence of preterm birth within 48 hours,<sup>41</sup> prevalence of preterm birth within 48 hours was lower (2.4%, 95% CI 0.1% to 12.9%) than in any of the Actim Partus studies discussed in the previous section. Sensitivity was 0.0% (95% CI 0.0% to 97.5%) and specificity was 97.5% (95% CI 86.8% to 99.9%); the total sample size was 41 and only one test result was positive (a false positive).<sup>41</sup> These data, along with calculated values for PPV, NPV, LR+, LR-, concordance and yield, are given in *Table 7*.

**TABLE 7** Calculated diagnostic accuracy parameters against the 48-hour delivery reference standard

Study (first author and year)	Participants (n)	Diagnostic accuracy (95% CI)								
		Sensitivity (%)	Specificity (%)	LR+	LR-	PPV (%)	NPV (%)	Prevalence (%)	Concordance	Yield
<b>Actim Partus</b>										
Brik (2010) <sup>48</sup>	276	73.9 (51.6 to 89.8)	64.8 (58.6 to 70.7)	2.10 (1.56 to 2.82)	0.40 (0.20 to 0.81)	16.0 (9.6 to 24.4)	96.5 (92.5 to 98.7)	8.3 (5.4 to 12.2)	0.66 (0.60 to 0.71)	0.38 (0.33 to 0.44)
Goyal (2016) <sup>52</sup>	60	65.7 (47.8 to 80.9)	56.0 (34.9 to 75.6)	1.49 (0.90 to 2.47)	0.61 (0.34 to 1.09)	67.6 (49.5 to 82.6)	53.8 (33.4 to 73.4)	58.3 (44.9 to 70.9)	0.62 (0.48 to 0.74)	0.57 (0.43 to 0.69)
Lembet (2002) <sup>53</sup>	36	93.3 (68.1 to 99.8)	81.0 (58.1 to 94.6)	4.90 (2.01 to 11.96)	0.08 (0.01 to 0.55)	77.8 (52.4 to 93.6)	94.4 (72.7 to 99.9)	41.7 (25.5 to 59.2)	0.86 (0.71 to 0.95)	0.50 (0.33 to 0.67)
Ting (2007) <sup>56</sup>	94	100.0 (47.8 to 100.0)	74.2 (63.8 to 82.9)	3.87 (2.72 to 5.50)	0.00 (N/A)	17.9 (6.1 to 36.9)	100.0 (94.6 to 100.0)	5.3 (1.7 to 12.0)	0.76 (0.66 to 0.84)	0.30 (0.21 to 0.40)
Tripathi (2016) <sup>57</sup>	468	95.5 (89.7 to 98.5)	82.1 (77.8 to 86.0)	5.34 (4.26 to 6.69)	0.06 (0.02 to 0.13)	62.1 (54.4 to 69.5)	98.3 (96.1 to 99.5)	23.5 (19.7 to 27.6)	0.85 (0.82 to 0.88)	0.36 (0.32 to 0.41)
Vishwekar (2017) <sup>58</sup>	30	84.6 (54.6 to 98.1)	82.4 (56.6 to 96.2)	4.79 (1.67 to 13.74)	0.19 (0.05 to 0.68)	78.6 (49.2 to 95.3)	87.5 (61.7 to 98.4)	43.3 (25.5 to 62.6)	0.83 (0.65 to 0.94)	0.47 (0.28 to 0.66)
<b>PartoSure</b>										
Werlen (2015) <sup>41</sup>	41	0.0 (0.0 to 97.5)	97.5 (86.8 to 99.9)	0.00 (N/A)	1.03 (0.98 to 1.08)	0.0 (0.0 to 97.5)	97.5 (86.8 to 99.9)	2.4 (0.1 to 12.9)	0.95 (0.84 to 0.99)	0.02 (0.00 to 0.13)
N/A, not applicable.										



**FIGURE 5** The ROC plot for Actim Partus against the 48-hour delivery reference standard.

## Summary

Diagnostic test accuracy data were sought in two ways:

1. a systematic review evaluating the test accuracy of: PartoSure, Actim Partus and qfFN at thresholds other than 50 ng/ml
2. 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 50 ng/ml (qualitative or quantitative tests).

## Data derived from the systematic review of diagnostic test accuracy

### 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<sup>42,43</sup>) included both Actim Partus and qfFN and the other (Hadzi-Lega *et al.*<sup>44</sup>) included both Actim Partus and PartoSure.
- Fourteen studies assessing only Actim Partus.<sup>45–58</sup>
- Three studies assessing only PartoSure.<sup>41,59–61</sup>
- One study assessing only qfFN.<sup>62</sup>

All 20 studies evaluated DTA against a reference standard of preterm delivery within 7 days.<sup>41–62</sup> In seven studies (six Actim Partus studies and one PartoSure study), test accuracy was also measured against a reference standard of preterm delivery within 48 hours.<sup>41,48,52,53,56–58</sup>

In the studies assessing two index tests in the same sample, APOSTEL-1<sup>42,43</sup> and Hadzi-Lega *et al.*,<sup>44</sup> sample sizes were 350 and 57, respectively. Sample sizes in the other studies ranged from 30 to 468 for Actim Partus<sup>57,58</sup> and from 41 to 203 for PartoSure,<sup>41,60,61</sup> and the only study evaluating qfFN alone comprised 455 participants.<sup>62</sup>

In addition, seven ongoing trials were identified that may be relevant to this review question, including four trials conducted in the UK (two of which aim to recruit > 1000 participants).

## Heterogeneity between studies

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

- The prevalence of preterm birth – prevalence of preterm delivery within 7 days ranged from 1.7% (95% CI 0.6% to 3.7%) to 73.3% (95% CI 60.3% to 83.9%) and prevalence of preterm delivery within 48 hours ranged from 2.4% (95% CI 0.1% to 12.9%) to 58.3% (95% CI 44.9% to 70.9%).<sup>41,49,52</sup>
- 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,<sup>42–44,59,62</sup> three further studies mentioned iatrogenic delivery as a reason for exclusion, but it is unclear how many of these deliveries took place within the time frame of the reference standard,<sup>48,58,60,61</sup> and three studies report the number of spontaneous/iatrogenic deliveries but include the data from these women.<sup>47,49,55</sup> In the remaining 10 studies, the mode of delivery was not clearly reported.<sup>41,45,46,50–54,56,57</sup>
- Gestation – the majority of included studies used 24 weeks as the lower limit for gestation at enrolment,<sup>41–43,45–52,54–56,62</sup> with the lower limit in the remaining six studies ranging from 20 to 28 weeks.<sup>44,53,57–61</sup> The upper limit for gestation varied more between studies, ranging from 32.6 to 37 weeks.<sup>50,55,57,58</sup> No studies reported test accuracy data stratified by gestation.
- 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.<sup>56</sup> All other studies reported uterine contractions as a necessary indicator of preterm labour;<sup>41–48,50–55,57–62</sup> however, there was variation in the rate of uterine contractions necessary for inclusion.<sup>42,43,45,50–55,57,62</sup> 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.<sup>1,42,43,49,58,62</sup> In these studies, the proportion of study participants with multifetal pregnancies ranged from 6% to 20%.<sup>42,43,49</sup>
- Risk status – only one study clearly reports the risk status of participants.<sup>59</sup> Heterogeneity of studies regarding the risk status of women is, therefore, unclear.
- Dilatation threshold and cervical length – all but four studies included a dilatation threshold for exclusion.<sup>49,52,53,58</sup> Typically, the threshold was > 3 cm or ≥ 3 cm, but two studies had a lower threshold (> 2 cm and ≥ 2 cm).<sup>46,54</sup> In two studies, all included women had a transvaginal cervical length measurement ≤ 30 mm.<sup>42,43,50</sup>
- Other more specific exclusion criteria also varied between studies (e.g. cervical cerclage, previous tocolytic treatment, recent sexual intercourse, vaginal bleeding and prior cervical examination).
- Participant characteristics also differed between studies. These differences included average maternal age, gestational age at presentation and history of preterm delivery.

## Administration of index tests

Studies generally followed manufacturers' guidance on how to administer index tests. Key differences in how the tests were administered include:

- Two studies used frozen samples in their analysis.<sup>42,43,49</sup> 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.<sup>55</sup>

- Two Actim Partus studies collected samples from the posterior fornix rather than from the external cervical os.<sup>42,43,56</sup>

### Provision of treatment

It should be noted that providing treatment (tocolytics and/or corticosteroids) may have an impact on the occurrence of the reference standard (i.e. whether or not preterm delivery takes place), and this would have an impact on the test accuracy data.

Whether or not 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 had an impact on the test accuracy results.

### 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.<sup>47,48,50,58,60,61</sup>
- 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 preterm birth within 48 hours or within 7 days).<sup>42,43,49</sup> There is also no clear evidence regarding the likely impact of longer-term storage on the tests.
- In eight studies, there was lack of clarity regarding whether or not index test assessors were blinded to the results of additional diagnostic tests (e.g. cervical length);<sup>42,43,50,51,54,56,57,60-62</sup> however, owing to the nature of the index tests, there is little scope for bias to exist in their interpretation.
- In two Actim Partus studies, samples were collected from the posterior fornix rather than the external cervical os.<sup>42,43,56</sup>
- The lack of clarity regarding the administration of tocolytics, particularly in reference to test results, precluded a thorough evaluation of the effect of treatment on test accuracy data.

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

Table 8 summarises the results from the systematic review of DTA for the three index tests (PartoSure, Actim Partus and qfFN).

**TABLE 8** Summary of evidence and relative accuracy against the 7-day reference standard

Index tests	Actim Partus	PartoSure	qfFN at		
			10 ng/ml	200 ng/ml	500 ng/ml
Actim Partus	NA				
PartoSure	No difference (Hadzi-Lega <i>et al.</i> , 2017 <sup>44</sup> )	NA			
qfFN at					
10 ng/ml	Sensitivity of fFN superior, specificity of Actim Partus superior (APOSTEL-1 <sup>42,43</sup> )	Indirect evidence only	NA		
200 ng/ml	No difference (APOSTEL-1 <sup>42,43</sup> )	Indirect evidence only		NA	
500 ng/ml	Sensitivity of Actim Partus superior, specificity of fFN superior (APOSTEL-1 <sup>42,43</sup> )	Indirect evidence only			NA

NA, not applicable.

As can be seen in *Table 8*, Actim Partus and PartoSure were assessed in the same sample in one study (Hadzi-Lega *et al.*<sup>44</sup>) and Actim Partus and qfFN were assessed in the same sample in one other study (APOSTEL-1<sup>42–44</sup>). No studies were identified that assessed PartoSure and qfFN in the same sample.

As well as being assessed in the same sample as Actim Partus in the study by Hadzi-Lega *et al.*,<sup>44</sup> PartoSure was assessed in the same sample as the QuikCheck test in one study (Nikolova *et al.*<sup>61</sup>).

### Summary of test accuracy data across the systematic review

*Table 9* summarises the sensitivity and specificity data for the index tests in the systematic review of test accuracy (PartoSure, Actim Partus and qfFN at thresholds other than 50 ng/ml).<sup>24</sup>

**TABLE 9** Summary of test accuracy data from the systematic review

Index test	Source	Test accuracy data (%) (95% CI)	
		Sensitivity	Specificity
<b>Test accuracy for the prediction of preterm delivery within 7 days</b>			
1. Studies assessing more than one index test			
fFN at 10 ng/ml	Bruijn <i>et al.</i> (APOSTEL-1) <sup>42,43</sup> (n = 350)	95.7 (87.8 to 99.1)	42.3 (36.5 to 48.4)
fFN at 200 ng/ml	Bruijn <i>et al.</i> (APOSTEL-1) <sup>42,43</sup> (n = 350)	71.0 (58.8 to 81.3)	83.6 (78.8 to 87.8)
fFN at 500 ng/ml	Bruijn <i>et al.</i> (APOSTEL-1) <sup>42,43</sup> (n = 350)	42.0 (30.2 to 54.5)	95.7 (92.7 to 97.8)
Actim Partus	Bruijn <i>et al.</i> (APOSTEL-1) <sup>42,43</sup> (n = 350)	78.3 (66.7 to 87.3)	89.3 (85.1 to 92.7)
PartoSure	Hadzi-Lega <i>et al.</i> (2017) <sup>44</sup> (n = 57)	83.3 (35.9 to 99.6)	90.2 (78.6 to 96.7)
Actim Partus	Hadzi-Lega <i>et al.</i> (2017) <sup>44</sup> (n = 57)	83.3 (35.9 to 99.6)	76.5 (62.5 to 87.2)
2. Studies assessing a single index test			
Actim Partus	Pooled (16 studies)	77.0 (68.0 to 83.0)	81.0 (76.0 to 85.0)
	Range (16 studies)	33.3 (4.3 to 77.7) to 94.7 (89.9 to 97.7)	50.0 (24.7 to 75.3) to 93.5 (82.1 to 98.6)
PartoSure	Pooled (4 studies)	83.0 (61.0 to 94.0)	95.0 (89.0 to 98.0)
	Range (4 studies)	0.0 (0.0 to 97.5) to 100.0 (73.5 to 100.0)	90.2 (78.6 to 96.7) to 97.5 (96.8 to 99.9)
fFN at 10 ng/ml	Range (2 studies)	93.8 (82.8 to 98.7) to 95.7 (87.8 to 99.1)	32.2 (27.7 to 37.0) to 42.3 (36.5 to 48.4)
fFN at 200 ng/ml	Range (2 studies)	70.8 (55.9 to 83.0) to 71.0 (58.8 to 81.3)	78.6 (74.3 to 82.5) to 83.6 (78.8 to 87.8)
fFN at 500 ng/ml	Range (2 studies)	29.2 (17.0 to 44.1) to 42.0 (30.2 to 54.5)	94.3 (91.6 to 96.4) to 95.7 (92.7 to 97.8)
<b>Test accuracy for the prediction of preterm delivery within 48 hours</b>			
3. Studies assessing a single index test			
Actim Partus	Pooled (6 studies)	87.0 (74.0 to 96.0)	73.0 (62.0 to 82.0)
	Range (6 studies)	65.7 (47.8 to 80.9) to 100 (47.8 to 100.0)	56.0 (34.9 to 75.6) to 82.4 (56.6 to 96.2)
PartoSure	Werlen <i>et al.</i> <sup>41</sup> (2015) (n = 41)	0.0 (0.0 to 97.5)	97.5 (86.8 to 99.9)

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

- Data set 1 (see *Table 9*) reports test accuracy data obtained from the two studies that assess two index tests (included in the systematic review).
- Data set 2 (see *Table 9*) 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 (see *Table 9*) 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 qFFN because only two studies of this test were included.

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

- Data set 3 (see *Table 9*) 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 48-hour reference standard. For Actim Partus, meta-analyses were conducted.



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

End-to-end studies investigate the clinical impact of conducting tests by following patients from testing, through treatment to final clinical outcomes. Randomised controlled trials (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 conducted a systematic review of end-to-end studies, with a particular focus on RCTs but also including other controlled study designs. This review was undertaken following the general principles published by the University of York Centre for Reviews and Dissemination.<sup>37</sup> The protocol was registered on PROSPERO (reference number CRD42017072696).

### Methods for reviewing effectiveness

#### Identification of studies

The same searches as for the review of diagnostic accuracy studies were conducted (see *Chapter 2, Identification of studies*). In brief, these included searches of electronic databases (these were designed to identify all studies assessing PartoSure, Actim Partus and qfFN), all systematic reviews identified by the electronic searches, trial registries, Google Advanced Search, reference lists of included DTA studies, studies citing the included DTA studies and industry submissions to NICE.

As with the review of DTA (see *Chapter 2, Methods of the systematic review*), 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 conducted 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). Prespecified inclusion and exclusion criteria were used (see *Inclusion and exclusion criteria*). Disagreements were resolved by discussion.

#### Inclusion and exclusion criteria

##### Population

Regarding the population, inclusion criteria were the same as those for the review of test accuracy studies; this includes the protocol amendment in which twin or multiple pregnancies could make up 20% of the total population (see *Chapter 2, Population*).

##### Interventions

The interventions under consideration were identical to those in the review of test accuracy (i.e. PartoSure, Actim Partus and qfFN; see *Chapter 2, Index tests*).

##### 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 qfFN 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.

## Outcomes

In accordance with the NICE scope,<sup>12</sup> 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 (i.e. they deliver within 7 days following treatment)
- number of women treated with maternal corticosteroids inappropriately (i.e. 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.

## Study design

Randomised controlled trials were primarily sought for this review; however, other controlled designs (prospective or retrospective) were also eligible for inclusion.

### *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.

## Results

### *Quantity and quality of research available*

After screening 2623 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 regarding the clinical outcomes of interest; there were no studies identified in which some women received one test and some received another and even in the studies identified in the test accuracy review, in which women received more than one test (see *Chapter 2, Summary*), 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.

### *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.

It is 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 provide the only available evidence.

## Summary

We were not able to draw any conclusions on the effectiveness of PartoSure, Actim Partus or qfFN 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 of 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 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). Therefore, it is worth considering whether or not conducting controlled studies in this area can reasonably be expected. On balance, it does not seem unreasonable to expect such studies to be conducted in this population; the principal barrier to conducting a RCT would be the potential difficulty of recruiting participants during an acute medical situation (e.g. the time needed to consent and randomise). However, this population and these tests would also lend themselves well to a RCT design with regards the length of follow-up required (for a number of key outcomes, the length of follow-up could be < 1 year).

Nevertheless, the decision to only include controlled studies was primarily based 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 provide the only available evidence. However, 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 available evidence) and these data may, therefore, be obtained from studies that have not been selected via a systematic reviewing process.



## Chapter 4 Data informing the economic modelling

The systematic review produced limited DTA data (see *Chapter 2*) and no clinical effectiveness data (see *Chapter 3*) for populating an economic evaluation of diagnostic tests of interest. There was no single DTA study that evaluated all three index tests, and only two studies (APOSTEL-1<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup>) 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 dilatation 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 *Chapter 2* are likely to be biased because 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,<sup>42,43</sup> which assessed both Actim Partus and fFN, and Hadzi-Lega *et al.*,<sup>44</sup> which assessed Actim Partus and PartoSure. In addition, we excluded studies that investigated laboratory-based enzyme-linked immunosorbent assay (ELISA) qualitative (at 50 ng/ml) fFN tests, as this technology is no longer in use; thus, meta-analysis of the remaining four studies of Actim Partus versus qualitative fFN<sup>42,43,51,56,57</sup> provided the DTA results used in the economic evaluation. Details are presented in *Chapter 6*.



## Chapter 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 of the diagnostic test interventions. Owing to the limited evidence on economic evaluations evaluating index tests, the review was extended to include economic evaluation studies of any test identified in our systematic search of cost-effectiveness studies. The second part of this chapter presents a review of the modelling structures used in previous evaluations of diagnostic tests for the diagnosis of preterm labour in symptomatic women with intact membranes, identified by the same systematic search.

### 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 qfFN. The review was undertaken following the general principles published by the University of York Centre for Reviews and Dissemination.<sup>37</sup> The protocol was registered on PROSPERO (reference number CRD42017072696).

#### Identification of studies

The methods followed those reported in *Chapter 2, Methods of the systematic review*, for study identification. Studies were screened by two reviewers (RMM and JVC).

#### Inclusion and exclusion criteria

The population, index test and reference standard matched those reported in *Chapter 2, Inclusion and exclusion criteria*, and the inclusion criteria include the protocol amendment that twin or multiple pregnancies could make up 20% of the total population. However, for the review of economic evaluations, the criteria for inclusion permitted studies that reported health-care costs of an index test without restriction in terms of the design of the effectiveness study. In the following sections, any reference to exclusions owing to study design mean economic study design (i.e. exclusion of studies that are not economic evaluations).

#### Data extraction strategy

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

#### Critical appraisal strategy

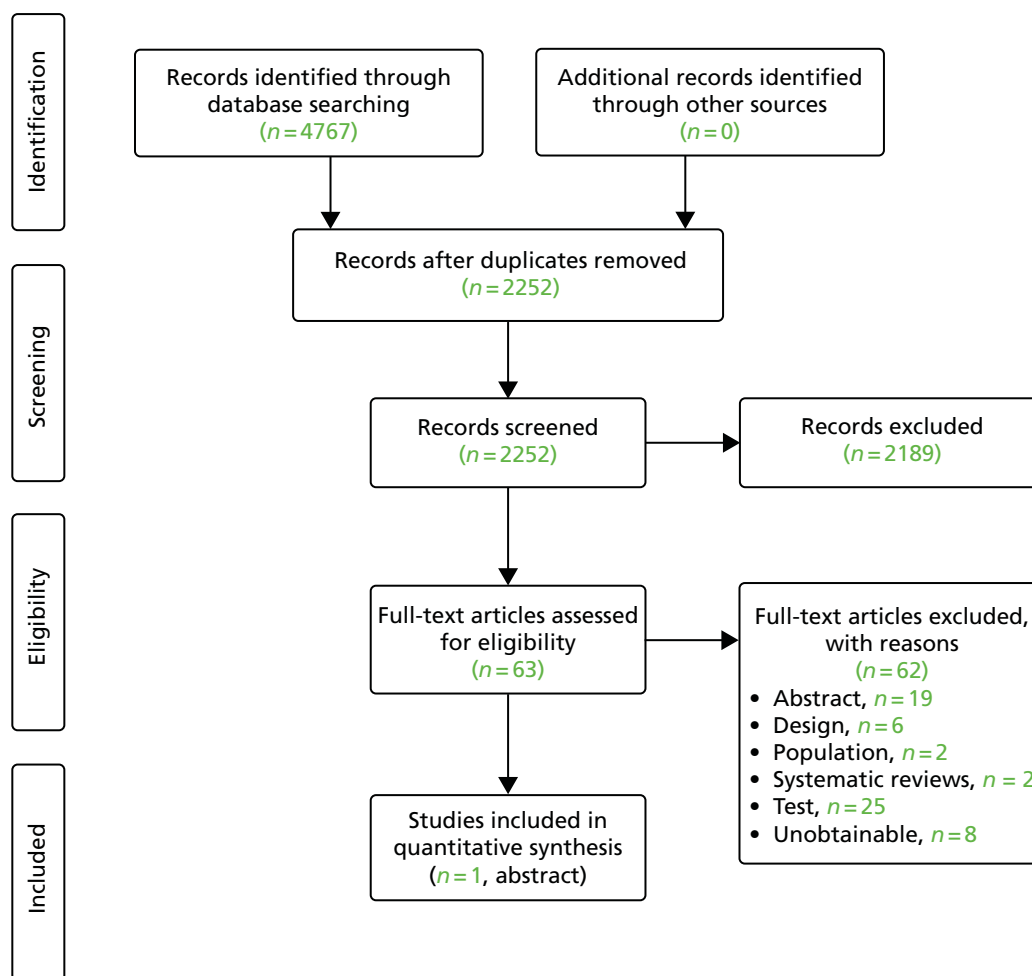
The quality of the studies was assessed in detail by an experienced health economist (RMM) in accordance with the criteria specified by the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist.<sup>64</sup> This represents a deviation from the study protocol, which stated that the Drummond and Jefferson's checklist would be used.<sup>65</sup>

#### Methods of data synthesis

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

## Results

Figure 6 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 2252 records, 63 citations were taken to full-text screening. One study met the inclusion criteria; it was a conference abstract<sup>67</sup> of a MSc dissertation.<sup>68</sup> We contacted the authors and they provided a copy of the dissertation, which is the basis for this review.



**FIGURE 6** The PRISMA flow diagram for the economic evaluation review.<sup>66</sup> © 2009 Moher *et al.*<sup>66</sup> This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### 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 ACSs and tocolysis with full compliance with the treatment protocol at different fFN thresholds (positive result of  $\geq 10$ ,  $\geq 50$ ,  $\geq 200$  and  $\geq 500$  ng/ml) provided by the Hologic, Inc. Rapid 10Q System against delivery outcomes. Clinicians were blinded to the qfFN concentration 'to prevent it influencing their management based on the qualitative fFN test result' (Table 10).<sup>67</sup>

The study reported the proportion of compliance with the qualitative fFN treatment protocol: 67% of positive cases (35/51) and 6% of negative cases (16/252) were given ACS 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% of women (10/75) testing positive with qualitative fFN were administered tocolytics, whereas 2% of women (6/282) testing negative received tocolytics. Of those patients who tested positive and were given tocolytics, 1 out of 10 delivered within 7 days and 4 out of 10 delivered before 37 weeks.



**TABLE 10** The included cost-effectiveness study

Study (first author and year)	Population	Setting	Test/diagnostic strategy	Study design	n	Time frame	Outcome and accuracy test results (if available)	Results	Comments
Gibson (2013) (dissertation) <sup>67,68</sup>	Subsample of EQUIPP: high-risk* symptomatic women aged $\geq 18$ years between 23 <sup>+0</sup> and 34 <sup>+6</sup> weeks' gestation (GA range suitable for ACS and tocolysis administration). Presenting symptom: TPTL (contractions > 2 in 30 minutes) 15.2%; abdominal pain 70.9% of women; remaining 13.9% presented with other symptoms, such as tightening and pelvic pressure	Preterm Surveillance Clinic at St Thomas' Hospital in London	a. Management without fFN testing (treat all) b. Qualitative fFN protocol in reality (less than full compliance) c. Full adherence to qualitative fFN protocol d. Modelled implementation of qfFN at 10-, 50-, 200- and 500-ng/ml thresholds e. Clinicians blinded to qfFN results. Aim was to evaluate whether or not qfFN could add value to clinical management protocol	Prospective cohort	ACS analyses: 306 Tocolysis 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 or not they received ACS, because of the evidence suggesting ACS is ineffective in the reduction of RDS when delivery occurs within 24 hours of treatment <sup>69</sup>  On the finding that the 500-ng/ml threshold results in 12 fewer women treated with ACS and one missed case in need of treatment relative to the 200-ng/ml threshold, whereas no missed cases occur at the 200-ng/ml 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

GA, gestational age; NNT, number needed to treat; PTB, preterm birth; RCOG, Royal College of Obstetricians and Gynaecologists; SPTB, spontaneous preterm birth; TPTL, threatened preterm labour.

a At least one of the following risk factors: history of previous PTB (< 37 weeks' gestation)/second trimester loss ( $\geq 16$  weeks' gestation); short cervical length (< 25 mm) measured on ultrasound scan at 18<sup>+0</sup>–27<sup>+6</sup> weeks' gestation; previous cervical surgery. Exclusion criteria: congenital abnormality; sexual intercourse within 24 hours; blood-stained swab; previously administered steroids/tocolysis; symptomatic visit number  $\geq 2$ .

In the published abstract,<sup>67</sup> 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: 1540 and 80, respectively. However, the methods used to obtain these numbers are not given in the abstract and no reference to these results appears in the dissertation.<sup>68</sup> However, the dissertation does provide 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 one woman delivering within 7 days of testing (no test and treat all group, 77; 200 ng/ml fFN group, 9) and women delivered before receiving a full steroids course (i.e. within 24 hours of testing; three in both cases).

**Critique**

Although the study by Gibson *et al.*<sup>67</sup> 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: (1) no test and treat all with steroids versus qualitative testing with fFN and treat those with positive results, and (2) testing options investigated at the qfFN thresholds of 10, 50, 200 and 500 ng/ml.

Given the available data from the study, the costs per patient adequately treated with steroids (i.e. within 7 days) were calculated by the assessment group (AG) using the following formula (for the comparison between no test and treating all with steroids):

$$\begin{aligned} &\text{Incremental cost per additional patient adequately treated} \\ &= \{C_{ACS} - [C_{fFN} + fFN_{+} \times (C_{ACS} + C_H) + fFN_{-} \times (C_{ACS} + C_H)]\} \frac{N}{N_{TA} - N_{fFN}} \quad (7) \\ &= \{C_{ACS} - [C_{fFN} + (fFN_{+} + fFN_{-}) \times (C_{ACS} + C_H)]\} \times NNT, \end{aligned}$$

Where  $C_{ACS}$  = cost of steroids,  $C_{fFN}$  = cost of fFN 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 and at least 1 day before delivery in the treat-all strategy,  $N_{fFN}$  = number of mothers given ACS within 7 days and at least 1 day before delivery in the testing strategy,  $fFN_{-}$  = probability of a false-negative test result and  $NNT$  = 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) multiplied by the incremental cost per patient of treating all versus testing and treating positive cases. By combining the data from Gibson *et al.*<sup>67</sup> with treatment effectiveness data and cost and utility values used by previous models (e.g. NICE 2015,<sup>24</sup> discussed below), 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 quality-adjusted life-year (QALY) gained. The resulting formulae are:

$$\text{Incremental cost per RDS case avoided} = \{C_{ACS} + C_H - [C_{fFN} + (fFN_{+} + fFN_{-}) \times (C_{ACS} + C_H)]\} \times \frac{NNT}{ARR_{RDS}}, \quad (8)$$

$$\text{Incremental cost per IVH case avoided} = \{C_{ACS} + C_H - [C_{fFN} + (fFN_{+} + fFN_{-}) \times (C_{ACS} + C_H)]\} \times \frac{NNT}{ARR_{IVH}}, \quad (9)$$

$$\text{Incremental cost per death avoided} = \{C_{ACS} + C_H - [C_{fFN} + (fFN_{+} + fFN_{-}) \times (C_{ACS} + C_H)]\} \times \frac{NNT}{ARR_{Death}}, \quad (10)$$

$$\begin{aligned} \text{Incremental cost per QALY gained} &= \{C_{ACS} + C_H - [C_{fFN} + (fFN_{+} + fFN_{-}) \times (C_{ACS} + C_H)]\} \times \\ &\frac{NNT}{ARR_{Death} \text{MaxQALY} + ARR_{RDS} D_{RDS} \text{Disu}_{RDS} + ARR_{IVH} D_{IVH} \text{Disu}_{IVH}}, \quad (11) \end{aligned}$$

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

By adopting the values in the NICE guideline model<sup>24</sup> discussed below and summarised in *Table 11*, one may calculate the relevant incremental cost-effectiveness measures using the diagnostic test results reported by Gibson *et al.*<sup>67</sup>

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

Gibson *et al.*<sup>67</sup> found that variation of the threshold from 10 to 50 to 200 ng/ml resulted in the same number (two) of false-negative cases of women who delivered within 7 days of testing. In terms of cost-effectiveness analysis, this finding means that qualitative testing using the 10 ng/ml and 50 ng/ml thresholds is dominated by testing at the 200 ng/ml threshold because the latter results in lower resource use for testing and hospital admissions than the lower threshold strategies. However, moving from the 200 ng/ml to the 500 ng/ml threshold resulted in one additional missed preterm birth case.<sup>67</sup> Comparing the qfFN test with the 200 versus the 500 ng/ml threshold, the AG calculates incremental costs per event avoided of £221,115, £770,000 and £235,714 for the RDS, IVH and death outcomes, respectively. 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 qfFN with a threshold of 200 ng/ml.

**TABLE 11** Model parameter values from the NICE guideline model<sup>24</sup>

Parameter	Parameter definition	Values	Source
$C_{ACS}$	Cost of full ACS course	Included in cost of hospital admission ( $C_H$ )	NICE guideline 2015 <sup>24</sup>
$C_{fFN}$	Cost of fFN test	£37.50	NICE guideline 2015 <sup>24</sup>
$C_H$	Cost of hospital admission	£1050	NICE guideline 2015 <sup>24</sup>
$fFN_+$	Marginal probability of positive fFN test result	0.18	Gibson <i>et al.</i> <sup>67</sup>
$fFN_-$	Marginal probability of false-negative fFN test result	0.0082	Gibson <i>et al.</i> <sup>67</sup>
NNT	Number needed to treat to avoid one inadequately treated case	28.5	Gibson <i>et al.</i> <sup>67</sup>
$ARR_{Death}$	Absolute risk reduction of death from treatment	0.049	NICE guideline 2015 <sup>24</sup>
$ARR_{RDS}$	Absolute risk reduction of RDS from treatment	0.052	NICE guideline 2015 <sup>24</sup>
$ARR_{IVH}$	Absolute risk reduction of IVH from treatment	0.015	NICE guideline 2015 <sup>24</sup>
$D_{RDS}$	Death risk from RDS	0.054	NICE guideline 2015 <sup>24</sup>
$D_{IVH}$	Death risk from IVH	0.300	NICE guideline 2015 <sup>24</sup>
MaxQALY	Maximum lifetime QALYs without RDS or IVH	22.44	NICE guideline 2015 <sup>24</sup>
$Disu_{RDS}$	QALY loss from RDS	3.85	NICE guideline 2015 <sup>24</sup>
$Disu_{IVH}$	QALY loss from IVH	4.5	NICE guideline 2015 <sup>24</sup>

It is evident from the small numbers of false-negative cases presented above that the findings from the study by Gibson *et al.*<sup>67</sup> are highly uncertain. This also highlights the need for evidence synthesis over multiple studies in order to derive meaningful evidence.

We highlight that the above formulae (equations 7–11) allow for a separate treatment of the costs of hospital admission and steroid treatment, in contrast to other models discussed below. This may be important because 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 qfFN thresholds.

### 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<sup>67</sup> and corresponding dissertation<sup>68</sup> 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 qfFN test. We note that current treatment protocols in some hospitals may allow for the use of different qfFN thresholds to decide whether or not to administer steroids and admit to hospital.

### Observational cost-minimisation studies

A set of studies was identified by one reviewer (RMM) in the systematic search of electronic bibliographic databases that investigated the health-care costs of the comparator in this review: qualitative fFN testing versus 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. For details of these studies, see *Appendix 4*.

### Model-based studies

We identified six different model structures presented in modelling studies of diagnostic interventions of preterm labour.<sup>70–75</sup> We describe these models as presented in their most recent applications found in the published literature (*Table 12*). 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 guidelines<sup>24</sup> by the Royal College of Obstetricians and Gynaecologists (RCOG). The second aspect is the length of analytical horizon; 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, with the model assuming 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<sup>24</sup> 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.

A detailed review of model-based studies can be found in *Appendix 4*. Here, we restrict discussion to a single study, the 2015 NICE guidelines model,<sup>24</sup> as this forms the basis for the independent economic assessment model we develop in *Chapter 6*.

**TABLE 12** Modelling studies identified in the systematic review of cost-effectiveness evidence

Study (first author and year)	Population	Perspective	Setting	Test/diagnostic strategy	Model structure	Time frame	Effectiveness and cost parameters	Type of study	Comments
Deshpande (2013) <sup>70</sup> Update of Honest (2009) <sup>71</sup>	Threatened PTL	NHS	Hospital	fFN and clinical examination vs. clinical examination alone	Decision tree	Before delivery	<ul style="list-style-type: none"> <li>• Steroids and tocolytics costs</li> <li>• Admission costs (LOS)</li> <li>• Hospital transfers</li> <li>• Ultrasound scan (after admission)</li> </ul>	Cost-minimisation	<ul style="list-style-type: none"> <li>• Documents the use of steroids and tocolytics from UK diagnostic study data reported by Dutta and Norman (2011)<sup>76</sup></li> <li>• Useful source of data on cost parameters</li> <li>• Does not include cost of delivery</li> </ul>
Chuck (2015) <sup>72</sup>	Threatened PTL or early onset of delivery in administrative databases (inpatient and outpatient)  Alberta, Canada	Health system	Hospital	fFN vs. no fFN	Decision tree	Delivery	<ul style="list-style-type: none"> <li>• Test specificity and sensitivity</li> <li>• Testing rates</li> <li>• Admission rates after testing</li> <li>• Transfers</li> <li>• LOS</li> <li>• Health-care costs</li> </ul>	Cost-minimisation	<ul style="list-style-type: none"> <li>• Populated with data from observational study (rates of admission and transfers)</li> <li>• The observational study found that fFN increased transfers and admissions</li> <li>• Assumed no costs savings from reductions of hospital admissions for episodes with false negative results, as costs would only be delayed</li> </ul>
Boyd (2011) <sup>73</sup>	Threatened PTL (clinical diagnosis)  ≥ 24 weeks' gestation  UK	NHS	Hospital	fFN vs. no test	Decision tree	3 months post birth or neonatal hospital discharge	<ul style="list-style-type: none"> <li>• Hospitalisations</li> <li>• Transfers</li> <li>• Health-care costs</li> <li>• Neonatal morbidity and mortality</li> </ul>	Cost-effectiveness	<ul style="list-style-type: none"> <li>• fFN saved costs but had a 'small but potentially detrimental' increase in neonatal morbidity, and a 'negligible increase in mortality'</li> <li>• It did not measure negative effects of steroids use in false positives. Only preterm infants subject to morbidity are exposed to mortality risk</li> </ul>
Mozurkewich (2000) <sup>74</sup>	Threatened PTL (regular uterine contractions), 24 to 34 weeks' gestation, intact membranes, without cervical dilatation ≥ 3 cm  USA	Third-party payer	Tertiary care unit	Rapid fFN vs. treat all with steroids as outputs	Decision tree	Neonatal hospital discharge or death	<ul style="list-style-type: none"> <li>• Total cost = triage or outpatient + fFN testing + hospitalisation and treatment + maternal delivery + neonatal care</li> <li>• RDS</li> <li>• Neonatal deaths</li> </ul>	<ul style="list-style-type: none"> <li>• Cost-effectiveness</li> <li>• Cost per case of RDS/ neonatal death avoided</li> </ul>	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 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

continued

**TABLE 12** Modelling studies identified in the systematic review of cost-effectiveness evidence (*continued*)

Study (first author and year)	Population	Perspective	Setting	Test/diagnostic strategy	Model structure	Time frame	Effectiveness and cost parameters	Type of study	Comments
NICE (2015) <sup>24</sup>	Women with suspected PTL and intact membranes  England	NHS	Hospital	Treat all vs. test and treat positive cases	Decision tree	Lifetime of infant	<ul style="list-style-type: none"> <li>• QALYs</li> <li>• Costs (tests, treatment, maternal and neonatal admissions, lifelong health care)</li> <li>• Sensitivity and specificity of SPTB within 7 days</li> <li>• Mortality</li> <li>• RDS</li> <li>• IVH</li> </ul>	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) <sup>75</sup>	Women with symptoms of preterm labour, <sup>a</sup> intact membranes and gestational ages of 24–34 weeks (APOSTEL-I)  The Netherlands	Societal	Hospital	Treat all vs. fFN	Decision tree	Neonatal death or hospital discharge	<ul style="list-style-type: none"> <li>• Proportion of patients treated, perinatal death, composite of adverse neonatal outcomes (perinatal death, CLD, neonatal sepsis, IVH &gt; grade II, periventricular leucomalacia &gt; grade I, and necrotising enterocolitis)</li> <li>• Costs (health care, medical and non-medical transport and indirect)</li> </ul>	<ul style="list-style-type: none"> <li>• Cost-effectiveness</li> <li>• Cost per death avoided</li> <li>• Cost per neonatal adverse event avoided</li> </ul>	<p>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 centre 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</p> <p>Accounted for different levels of intensity of care (i.e. admission on medium intensity, high intensity or intensive care wards) and in-utero transfers</p>

CLD, chronic lung disease; LOS, length of stay; PTL, preterm labour; SPTB, spontaneous preterm birth.  
 a Contractions (more than three in 30 minutes), vaginal bleeding or abdominal or back pain.

### *The National Institute for Health and Care Excellence 2015 guidelines model*

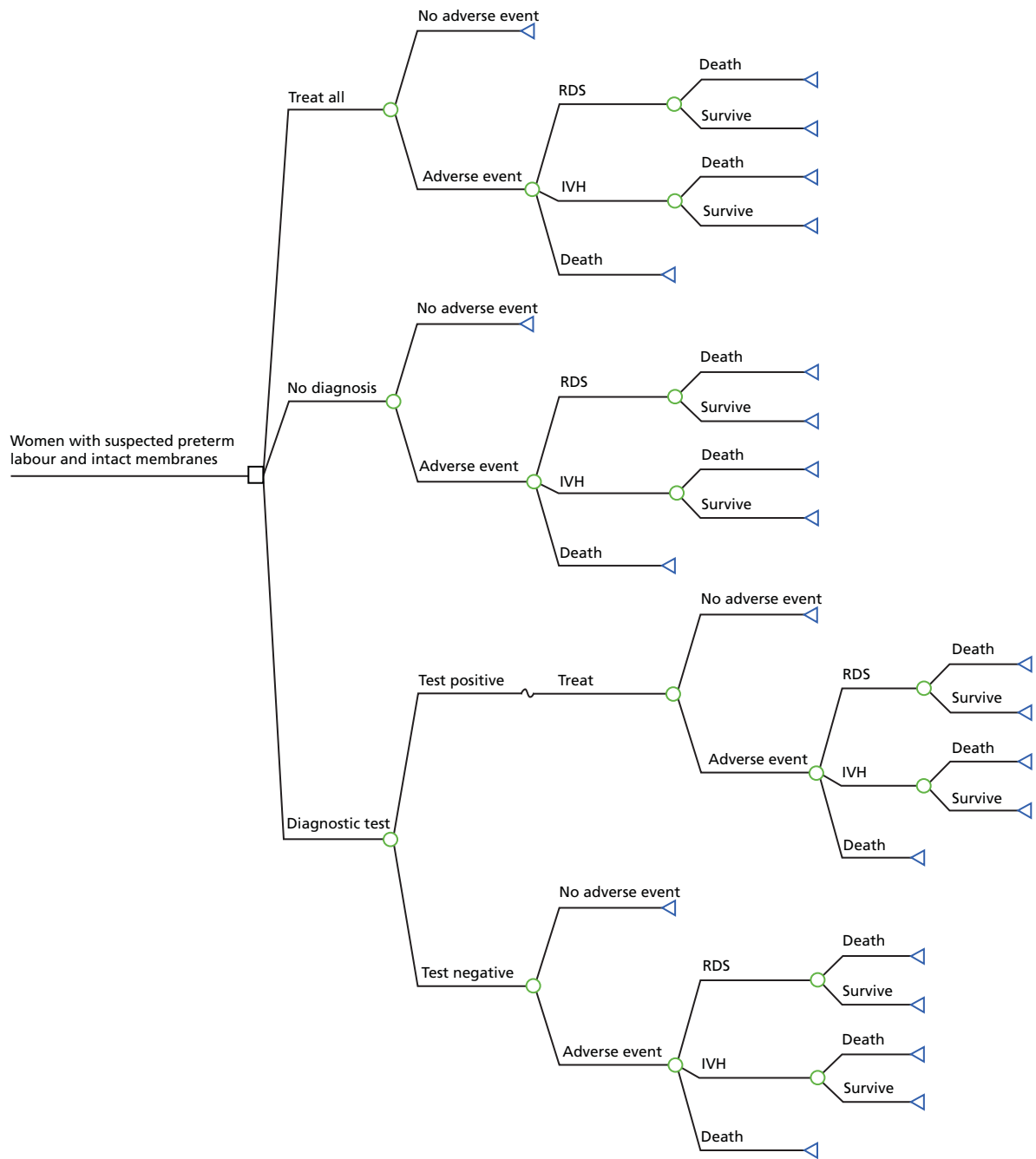
The RCOG developed a decision-analytic model to inform the NICE 2015 guidelines<sup>24</sup> on the diagnosis and treatment of preterm labour and birth. The authors of this model concluded that the quality of the diagnostic accuracy data was low for the different tests considered relevant at the time for women presenting with symptoms of preterm labour (cervical length measurement by ultrasound scan, Actim Partus, qualitative fFN). Consequently, they presented a 'what-if' analysis comparing the testing versus the no testing and treat-all strategies, which consisted of identifying the levels of specificity and sensitivity at which a hypothetical test became cost-effective in accordance with 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 at gestations of < 30 weeks. This served as the basis of the NICE recommendations about the use of testing to rule out preterm labour.

The NICE guidelines model structure is illustrated in *Figure 7*. In accordance with this model, the causal pathway from diagnostic results to neonatal outcomes is mediated by tocolysis treatment, which can delay premature delivery by  $\geq 48$  hours. This would generate a window of opportunity for appropriate steroid administration (i.e.  $\geq 24$  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 versus placebo, from three separate network meta-analyses of RCTs (one per model outcome) (NICE 2015).<sup>24</sup> 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 the OR 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 gestation in the model (24 to 34 weeks), so that the absolute risk reduction (ARR = relative risk  $\times$  baseline risk) for the three types of event with tocolysis varies by gestation only because the baseline (i.e. without tocolysis) risk declines with gestational age (*Figure 8* and see *Table 11*). IVH and RDS each contribute to the risk of neonatal mortality; the probability of death is 0.300 conditional on the former and 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.

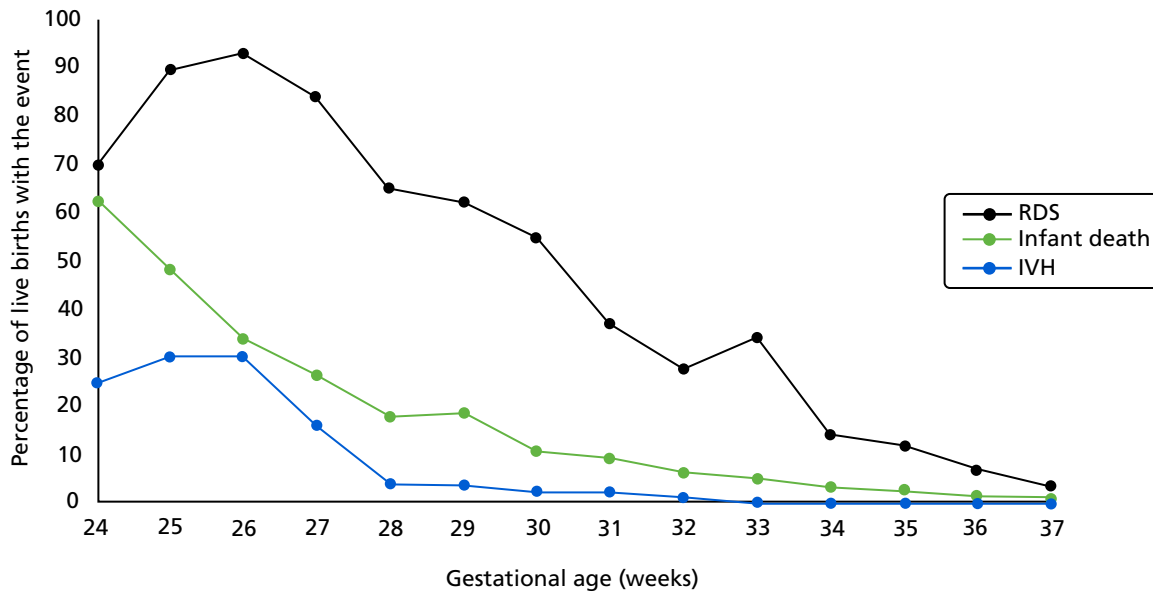
The analysis projected lifelong QALY values based on the neonatal adverse events. Infants who survived the neonatal phase and were discharged home were assumed to have average expected QALYs that varied with gestational age from 19.92 at 24 weeks to 22.61 at 34 weeks. These were calculated as the gestational age-specific proportion of infants surviving the first year of life multiplied by 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%.<sup>77</sup> Deaths in the first year of life were assumed to generate zero QALYs. In the event of RDS, an average QALY loss would apply, arbitrarily set at 3.85 (based on 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 the cerebral palsy value reported by Cahill *et al.*<sup>78</sup>

The analysis accounted for the costs of tests, including test acquisition and staff time, drug treatment [nifedipine at a loading dose of 40 mg and subsequent dose of 240 mg, at *British National Formulary* (BNF) prices of £0.008 per 1 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 neonatal intensive care unit (NICU) care [British Association of Perinatal Medicine (BAPM) level 1] with extracorporeal life support/extracorporeal



**FIGURE 7** The NICE 2015 guidelines model.<sup>24</sup> Reproduced from: Royal College of Obstetricians NICE Guideline 25 Preterm Labour and Birth, London, RCOG, November 2015, with the permission of the Royal College of Obstetricians and Gynaecologists.<sup>24</sup>





**FIGURE 8** Baseline (without tocolytics) risks in the NICE guideline model. Reproduced from: Royal College of Obstetricians NICE Guideline 25 Preterm Labour and Birth, London, ROCG, November 2015, with the permission of the Royal College of Obstetricians and Gynaecologists.<sup>24</sup>

membrane oxygenation, and the costs of IVH were assumed to be equal to the lifetime health-care costs of ICH, which, in turn, were assumed to be equal to the health-care costs of severity grade III or grade IV cerebral palsy.<sup>79</sup> Regardless of how valid these clinical assumptions are, the cost of IVH appears to be underestimated, 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<sup>80</sup> 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 [Great British pounds] using an exchange rate of £0.83 = €1 . . . It was then converted into 2011/12 prices using the HCHS (Hospital and Community Health Service) Index. One study<sup>81</sup> suggested that 30% of ICH is of severity Grade III and Grade IV and therefore the cost of ICH was estimated as  $0.3 \times £79,000$ .*

*© NICE 2015.<sup>24</sup> Preterm Labour and Birth. Available from [www.nice.org.uk/guidance/ng25](http://www.nice.org.uk/guidance/ng25). All rights reserved. Subject to Notice of rights. NICE guidance is prepared for the National Health Service in England. All NICE guidance is subject to regular review and may be updated or withdrawn. NICE accepts no responsibility for the use of its content in this product/publication*

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 2011/12 (IVH) and 2012/13 (RDS) prices. The long-term adverse event cost of IVH was derived from a European study that reported results in 2000 euros,<sup>80</sup> converted to GBP (Great British pounds) using the exchange rate of £0.83 = €1.00 (the year this rate applied to was not provided), and reflated to 2011/12 prices.

The key assumptions of the NICE guidelines model are summarised in *Table 13*. 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 happen in routine practice (Professor Andrew Shennan, King's College London, 2017, personal communication). The model does not explicitly account for the use and effect of corticosteroids, only implicitly within the treatment

**TABLE 13** Key assumptions in the NICE 2015 guidelines model<sup>24</sup>

Assumption	Description	Critique
The choice of diagnostic strategy has no clinically and economically 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
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, <sup>82</sup> and Hogg, Penney and Carmichael, 2007 <sup>83</sup> )
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 (Professor Andrew Shennan, King's College London, 2017, personal communication)
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 (1) allow in-utero transfers and/or (2) treat with steroids	Some protocols on the use of qfFN (e.g. London's Guy's and St Thomas' NHS Foundation Trust) provide different guidelines for the decision to admit and the decision to treat with tocolytics and to treat with steroids
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
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 model parameters for the treatment effects on these outcomes is based entirely on indirect comparisons, and the treatment effect estimates are consistent with no effect (i.e. 95% OR credible intervals cross the value of 1)
Neonatal mortality may occur through the risk of death associated with RDS or IVH or background risks that decline with gestational age	The effect of tocolysis on neonatal mortality is divided between an indirect effect, operated through its effect on RDS and IVH, and a direct effect through other causes	The network meta-analysis evidence used to populate the model parameters for the treatment effects on this outcome was based on no head-to-head data for the comparison of tocolysis vs. no treatment, and the treatment effect estimate is consistent with no effect (i.e. 95% OR credible interval crosses the value of 1)
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 preterm birth
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 preterm birth

effect estimates obtained from the network meta-analysis of tocolysis studies discussed previously in this section. This feature makes the model less suitable for obtaining generalisable results for situations in which the corticosteroids are used without tocolytic therapies. A major limitation of the analysis is the high degree of uncertainty associated with the calculation of costs and QALYs, which were extrapolated to lifetime values 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) because it did not account for the survival curve profile in population life tables and the varying utility with age,<sup>84</sup> and the utility norms were derived from a study that predates the time that EuroQol-5 Dimensions (EQ-5D) scores were developed. It is unclear if 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 gestation at presentation make this model the most relevant among those available for guiding clinical decisions on individual patients, because other models did not produce results by length of gestation. A summary of the main features of the model is provided using the CHEERS checklist<sup>64</sup> in *Table 14*.

**TABLE 14** The CHEERS checklist for the NICE guidelines model<sup>64</sup>

Item	Item number	Recommendation	Reported details
<b>Methods</b>			
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 preterm labour  (Section 9.6, pp. 176–7, and Section 1.3, p. 350, of NICE 2015 guideline <sup>24</sup> )
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 present 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. Health-care costs are based on NHS Reference Cost sources <sup>79</sup> and costs of medications are from BNF prices <sup>85</sup>
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 gestational 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 the low quality of DTA 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 lifetime horizon is appropriate because 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 <sup>24</sup>
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

continued

**TABLE 14** The CHEERS checklist for the NICE guidelines model<sup>64</sup> (continued)

Item	Item number	Recommendation	Reported details
Measurement of effectiveness	8	Describe fully the methods used for the identification of included studies and synthesis of clinical data	<p>The NICE guideline model<sup>24</sup> was systematically searched for studies of test accuracy of biochemical test, cervical length measurement by ultrasound scan and clinical examination. However, it found that the identified studies were of low quality</p> <p>The NICE guideline analysis<sup>24</sup> updated a systematic review comparing tocolytic treatment classes using network meta-analysis. This method allowed the comparison of studies that were not investigated directly in any RCT, thus expanding the evidence base for informing the analysis</p>
Measurement and valuation of preference-based outcomes	9	If applicable, describe the population and methods used to elicit preferences for outcomes	<p>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 one-third of the utility loss from cerebral palsy</p>
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	<p>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<sup>79</sup> of NICU care (BAPM level 1), and the costs of IVH were assumed to be equal to the lifetime health-care costs of ICH, which in turn were assumed to be equal to the health-care costs of severity grade III or IV cerebral palsy. The calculations used in the model appear to underestimate the long-term costs of IVH</p>
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	<p>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 2011/12 (IVH) and 2012/13 (RDS) prices. The long-term adverse event cost (IVH) was derived from a European study<sup>80</sup> that reported results in Euros of 2010, converted to GBP using an exchange rate of 2014 and reflatd them to 2011/12</p>

**TABLE 14** The CHEERS checklist for the NICE guidelines model<sup>64</sup> (*continued*)

Item	Item number	Recommendation	Reported details
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 <sup>24</sup> used a decision tree model, with long-term QALY and costs pay-offs. This is reasonable given the limited number of neonatal outcome data on which to base modelling of medium- to long-term outcomes
Assumptions	13	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 (1) full adherence to the diagnostic protocol, (2) that the effects of diagnostic testing on neonatal outcomes are mediated through treatment with tocolytics, (3) that steroid use is not explicitly modelled but implicit in the tocolytic treatment effect values estimated from the literature, (4) that the relative effects of tocolytics are constant across gestational ages, (5) that neonatal morbidity outcomes are measured in terms of RDS and IVH and (6) conditional on surviving the first year of life and neonatal morbidity outcomes, lifetime QALYs are independent of gestational age at birth
Analytic methods	14	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

## 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 (IVH and RDS) 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*<sup>79</sup> of critical care BAPM levels 1–4 is low, as returns are based on the 2011 Healthcare Resource Group (HRG) definition, which does not correspond with BAPM nursing requirements (Eleri Adams, Oxford University Hospitals NHS Foundation Trust, 2017, personal communication). In 2016, a new HRG classification was introduced to be more in line with BAPM requirements but this is not yet being used for *Reference Cost*<sup>79</sup> returns. In view of the lack of reliable data from *NHS Reference Costs*,<sup>79</sup> we propose that NHS tariffs for the four levels of care may be the best available estimates of true economic cost. Furthermore, the costs of IVH were based on inappropriate calculations and data for another disease, and systematic searching of the literature for better estimates of this cost parameter seems worthwhile.

Other areas of uncertainty that deserve to be explored include:

- the mortality benefits of reducing the rate of false-negative cases for patients, who, depending on the nature of the local hospital (i.e. a level 2/3 vs. a level 1 hospital), may be at increased risk of mortality if birth takes place before 32 weeks' gestation<sup>86</sup>
- accounting for differences in treatment costs of positive cases in accordance with the level of hospital of presentation, owing 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 *Chapter 6*, we undertake these revisions to the NICE guideline model and populate them with the DTA evidence from the systematic review in *Chapter 2*.

## Chapter 6 Independent economic assessment

### 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.

### Model structure

In common with all previous studies (see *Chapter 5*), 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,<sup>24</sup> 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. Although our protocol stated the plan to use a decision tree model for the diagnostic phase and another decision tree for the treatment phase, the available data did not allow us to populate the decision tree for the treatment phase. Therefore, our model is a decision tree with lifetime costs and QALY pay-offs. 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 QALYs based on neonatal morbidity and mortality outcomes. The main features of our model are that it:

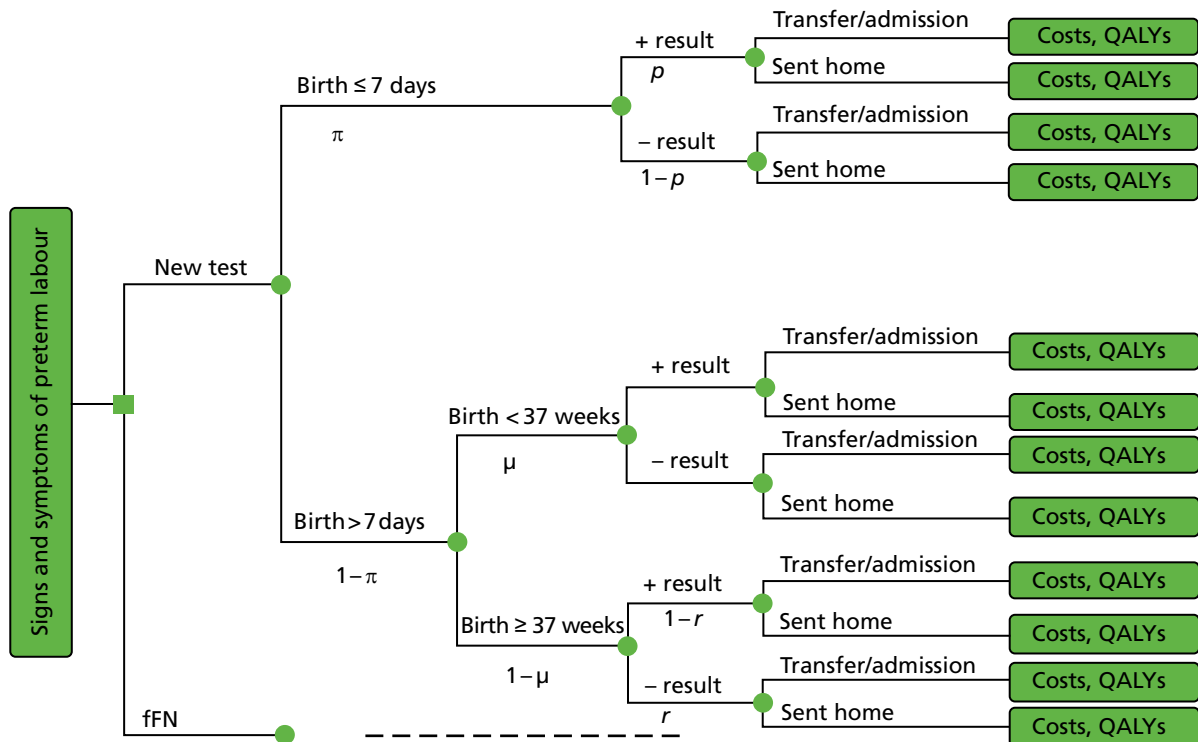
- accounts for the costs and QALYs of the infant (as well as QALYs for the mother, in a scenario analysis)
- differentiates costs and benefits by gestational age
- distinguishes between hospital levels of designation – level 3 hospitals have NICUs providing all types of care to the local population and care for the most severe infants transferred in utero or after birth, level 2 hospitals or local neonatal units provide all types of neonatal care except long-term intensive care and care for complex cases and level 1 hospitals provide specialised care for their local population, stabilisation and non-invasive monitoring
- 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<sup>24</sup> 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. 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 earlier in gestation. Unlike the NICE model, which assumed that the diagnosis of preterm labour was intended to guide the decision of whether or not to administer tocolysis, we model the treatment pathway following a diagnosis of preterm labour around the decision of whether or not 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 (ACS) administration for accelerating the maturation of the fetus's 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 infant and side-effects to the mother.<sup>30,33,34,87</sup> For example, audit data for the period from September 2016 to May 2017 from the level 2 hospital in Exeter show that tocolysis was administered in only one out of nine patients (11%) presenting with symptoms of preterm labour at 24–34 weeks with fFN of  $\geq 50$  ng/ml. In our analysis, we assume that tocolysis is used only for all in utero transfers at gestational ages of  $< 28$  weeks (see *Results*).

Given the importance of the timing of ACS administration, the aim was to develop a model capable of accounting for the different diagnostic test options' capabilities to distinguish between those women 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 days before delivery, few studies of DTA reported outcomes for the time-to-delivery interval of < 2 days (see *Chapter 2, Results of the systematic review*). These intervals have been discussed in the literature as most relevant to ACS effectiveness, with 7 days before delivery considered the earliest time for effective use of steroids in terms of fetal and neonatal mortality and RDS.<sup>88,89</sup> The latest evidence suggests that the effectiveness of ACSs in terms of mortality risk reduction may be optimal within 2 days of delivery, and that it diminishes with time before delivery. This observation also applies to the risk of RDS and IVH.<sup>33,87</sup> Thus, we decided on the model structure illustrated in *Figure 9*.

This structure shares the features of previous models of diagnosis and treatment and is determined by (1) the available DTA data and (2) the latest evidence on the time window relative to delivery when ACS 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 in accordance with their ability to correctly predict whether or not a woman will deliver before term, and whether or not that will happen before or after 7 days.

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 9* shows that a new test (PartoSure, Actim Partus and qfFN 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



**FIGURE 9** The Peninsula Technology Assessment Group cost-effectiveness model structure of diagnosis and treatment of women with symptoms of preterm labour. The dotted line indicates that the subsequent branch structure after 'fFN' is identical to the structure shown for 'New test'. New test is patient management in accordance with one of the interventions or 'index tests'. fFN is the comparator (admit and treat when fFN is  $\geq 50$  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.  $\pi$ , pre-test probability of preterm birth in  $\leq 7$  days;  $p$ , test sensitivity of PTB in  $\leq 7$  days;  $\mu$ , pre-test probability of PTB;  $r$ , test specificity of PTB. PTB, preterm birth.



(or qfFN used qualitatively at a 50 ng/ml threshold). The starting point of the model is when the decision between diagnostic strategies is made, that is, 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 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 happens, birth may be preterm (before 37 weeks' gestation) or full term. Therefore, women may be classified in one of these three subgroups in accordance with 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 ACS 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 < 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 that tocolysis is considered only for women undergoing in utero transfer at < 28 weeks' gestation) and transferred to a level 3 hospital for her care. Women who test negative are sent home without treatment; owing to a lack of any test-specific data on this parameter, we do not allow for partial compliance with treatment guidelines in contrast with 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.

In accordance with the model in *Figure 9*, a symptomatic woman who goes on to deliver within 7 days has a positive test result with probability  $\rho$  (the sensitivity of the test) and a negative test result with probability  $1 - \rho$  (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 ACS produces no benefit when administered > 7 days before preterm birth.<sup>88,89</sup> In scenario analyses, these women are assumed to benefit from ACS, but less so than those who are treated within 7 days of preterm delivery<sup>33,87</sup> (we assume throughout that no multiple courses of steroids are given, based on obstetricians' advice on routine practice and the perceived lack of proven benefit and risks to the neonate). The frequency of such cases is calculated as:

$$P(\text{PTB} > 7 \text{ days and +ve result}) = (1 - \rho) \times (1 - \text{Specificity}_{7d}) - (1 - \rho_{37w}) \times (1 - r), \quad (12)$$

where  $P(\text{PTB} > 7 \text{ days and +ve result})$  is the probability of having a positive test result and delivering preterm > 7 days after testing,  $\rho$  is the incidence of delivery within 7 days,  $\text{Specificity}_{7d}$  is the test specificity for delivery within 7 days and  $r$  is the test specificity for delivery at < 37 weeks (i.e. the proportion of women testing negative among those who deliver after 37 weeks), which has an incidence of  $\rho_{37w}$ . Thus the proportion of women who receive treatment more than 7 days before preterm birth (and, therefore, derive partial benefit from ACSs) is equal to the difference between the FPR for delivery within  $\leq 7$  days and the FPR for delivery is < 37 weeks, weighted by their respective incidences. The benefit from ACSs 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 a higher gestational age, than for mothers who deliver within 7 days. We assume that delivery takes place at the midpoint between gestation at presentation and 36 weeks.

### Population

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

### **Interventions and comparators**

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

- testing with PartoSure
- testing with Actim Partus
- testing with qfFN at thresholds of 10, 200 and 500 ng/ml
- comparator testing with fFN at 50 ng/ml, from quantitative or qualitative versions of the test device
- treat all without testing (i.e. because clinical investigation could not attribute symptoms to other causes, women are managed as presumptive cases 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 (1) qualitative fFN in accordance with 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 (2) qfFN (qfFN) as observed at the level 2 maternity hospital in Exeter.<sup>90</sup> These additional analyses are intended to reflect the spectrum of variation in local current practice across the country.

### **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, and the projected long-term health-care costs and utilities associated with avoiding an adverse neonatal outcome (death, RDS and IVH) were measured.<sup>91</sup> 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<sup>24</sup> (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.

### **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 ACS 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.

## **Treatment effectiveness and extrapolation**

### **Diagnostic test accuracy**

We limit our analyses to evaluate the diagnostic tests assessed by individual comparative diagnostic accuracy studies, as identified in *Chapter 2*. These studies are:

- APOSTEL-1, comparing Actim Partus with qualitative fFN at the 10, 50, 200 and 500 ng/ml thresholds (Bruijn *et al.*<sup>42,43</sup>)
- a comparison of Actim Partus with PartoSure (Hadzi-Lega *et al.*<sup>44</sup>).

Other studies were identified as providing relevant data but they were of lower quality than the data from these two studies. One was a comparison of Actim Partus with fFN at 50 ng/ml.<sup>49</sup> The study by Cooper *et al.*<sup>49</sup> was the second largest of the three identified studies that compared Actim Partus with fFN at 50 ng/ml and reported test accuracy data for delivery within 7 days and at < 37 weeks (see *Chapter 2, Results of the systematic review*). However, as discussed in *Chapter 2*, it was unclear what version of the fFN test was used, but the ELISA version of the qualitative fFN test, which is no longer used in clinical practice, was presumed to be used. 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.*<sup>44</sup> In scenario analyses, we also evaluated the comparison of Actim Partus with 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.<sup>42,43,51,56,57</sup> We did not consider this meta-analysis in the base case owing to the heterogeneity between the combined studies, especially in terms of preterm birth rates, which probably drove their differences in test accuracy results. Furthermore, the pooled results for each index test presented in *Chapter 2, Results of quantitative data synthesis (test accuracy data)*, were not considered in the economic analysis, because comparisons between tests in terms of those results are probably confounded by the heterogeneity between the studies.

In addition, although it was excluded from the assessment of test accuracy section (see *Chapter 2*) 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.*<sup>92</sup>) using data provided by the study authors for this review (Professor Andrew Shennan, King's College London, 2017, personal communication). This test was an assessment of the Rapid fFN 10Q analyser (Hologic, Inc.) at the qualitative thresholds of 10, 50, 200 and 500 ng/ml.

*Table 15* summarises the diagnostic test options compared in cost-effectiveness analyses, and the diagnostic accuracy study sources used to populate the economic model.

The sensitivities and specificities used for these analyses are presented in *Table 16*. Two sets of accuracy parameter values for each study were required for the model, one for predicting delivery within 7 days and another predicting 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.*<sup>42,43</sup>). For this study, therefore, we imputed specificity at 37 weeks from the corresponding 7-day specificity rate so as to obtain a 37-week FPR aligned with the UK study of fFN and the Italian study of Actim Partus.<sup>54,92</sup> Similarly, we imputed 37-week sensitivity values for the analysis of Actim Partus versus fFN (50-ng/ml threshold) based on our meta-analysis of 7-day accuracy data, using the Italian study data.<sup>54</sup> We varied these values in sensitivity analyses.

**TABLE 15** Diagnostic test options compared in cost-effectiveness analyses and their sources

Diagnostic test option	Comparator/ intervention	Base-case analysis		Scenario analysis		
		APOSTEL-1 <sup>42,43</sup>	Hadzi-Lega <i>et al.</i> <sup>44</sup>	Cooper <i>et al.</i> <sup>49</sup>	Abbott <i>et al.</i> <sup>92</sup>	Meta-analysis <sup>42,43,51,56,57</sup>
PartoSure	Intervention		✓			
Actim Partus	Intervention	✓	✓	✓		✓
Rapid fFN 10Q Cassette Kit thresholds other than 50 ng/ml	Intervention	✓			✓	
fFN, threshold of 50 ng/ml	Comparator	✓		✓	✓	✓
No test, treat all	Intervention	✓	✓	✓	✓	✓



Study (first author and year)	Delivery within 7 days				Probability distribution sensitivity [Beta( $\alpha$ , $\beta$ )]	Probability distribution specificity [Beta( $\alpha$ , $\beta$ )]	Delivery in < 37 weeks
	Diagnostic test	N	Sensitivity	Specificity			Specificity
Meta-analysis by AG <sup>42,43,51,56,57</sup>	Actim Partus	963	0.832	0.879	Beta(150,30)	Beta(689,94)	0.920 <sup>e</sup>
	fFN at 50 ng/ml	963	0.683	0.872	Beta(123,57)	Beta(683,100)	0.909 <sup>e</sup>

a Assumption: the FPR for delivery before 37 weeks was 6% lower than the FPR for delivery within 7 days of testing.  
b Assumption: the FPR for delivery before 37 weeks was 11% lower than the FPR for delivery within 7 days of testing.  
c Assumption: the FPR for delivery before 37 weeks was 18% lower than the FPR for delivery within 7 days of testing.  
d Assumption: the FPR for delivery before 37 weeks was 35% lower than the FPR for delivery within 7 days of testing.  
e 7-day FPR of test multiplied by ratio of FPR for delivery before 37 weeks relative to FPR for delivery in  $\leq 7$  days of corresponding test in Riboni *et al.*<sup>54</sup>  
f Imputed: 7-day FPR of test multiplied by ratio of FPR for delivery before 37 weeks relative to FPR for delivery in  $\leq 7$  days of corresponding test in Cooper *et al.*<sup>49</sup>

**Note**  
Further details are provided in *Chapter 2*, *The 7-day delivery reference standard*, and *Table 6*.

Figure 10 depicts the differences in diagnostic accuracy parameters for predicting delivery within 7 days across individual studies in Table 16 by index test and, for qfFN, by fFN threshold. The results from the study by Cooper *et al.*,<sup>49</sup> 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.<sup>49</sup> Because 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.<sup>44</sup>

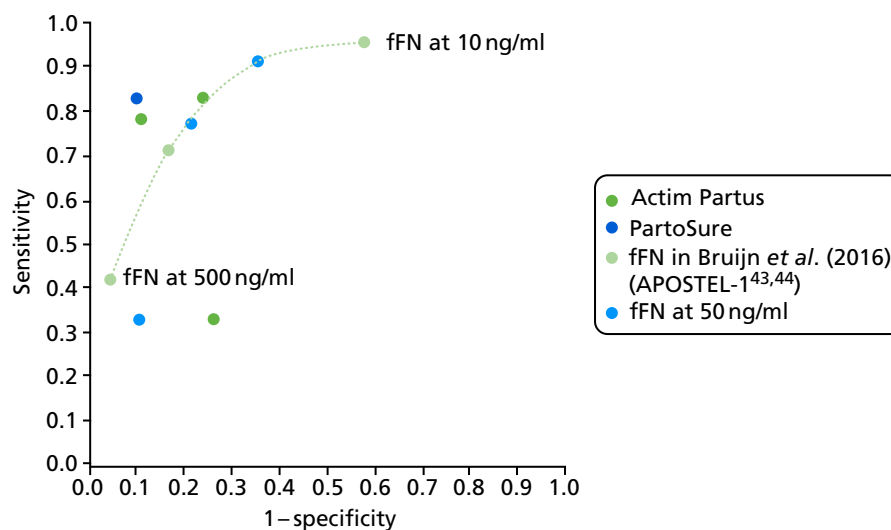
### Background neonatal risks parameter values

Our model included underlying risks of neonatal mortality and adverse events in terms of RDS and IVH, similar to the model that informed the NICE 2015 guidelines<sup>24</sup> on preterm labour diagnosis and treatment, but with data adjusted for steroid use in routine practice (Table 17 and Figure 11). We use the latest estimates of baseline mortality risks by gestational age from the ONS.<sup>13</sup> The risks of RDS and IVH were derived from Medscape data compiled by Michael Ross (author; MD, MPH Distinguished Professor of Obstetrics and Gynecology, University of California, Los Angeles, David Geffen School of Medicine; Distinguished Professor, Department of Community Health Sciences, Fielding School of Public Health at University of California at Los Angeles), which are available at <https://emedicine.medscape.com/article/260998-overview#a5> (accessed November 2017).

The baseline risk values in the model are intended to measure neonatal risks in the absence of ACS treatment; thus, the values in Table 17 have been adjusted to subtract the effect of steroids use in routine practice, using the formula:

$$\text{Baseline risk} = \frac{\text{Unadjusted risk}}{1 + P_{\text{ANS}} \times (\text{RR}_{\text{ANS}} - 1)} \quad (13)$$

where 'Unadjusted risk' is the risk estimate as reported in the data source,  $P_{\text{ANS}}$  is the prevalence of ACS use in routine practice,  $\text{RR}_{\text{ANS}}$  is the relative risk of mothers given ACSs relative to those not given ACSs 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 ACSs, in which the weights are given by the proportion of women receiving and not receiving ACSs. At the lowest extreme, the baseline risk will be



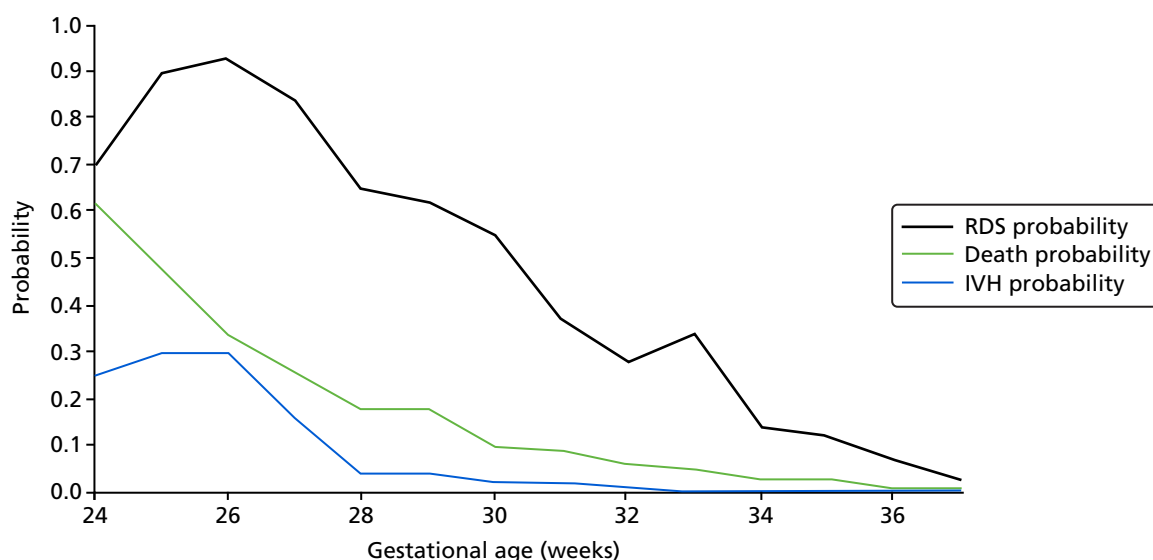
**FIGURE 10** Empirical summary ROC points across evaluated studies (7-day). These data (apart from those from Abbot *et al.*<sup>92</sup>) come from test accuracy studies assessed in Chapter 2 and are therefore subject to the issues discussed in relation to the quality of evidence (the implications of the quality of these data for the economic conclusions are discussed in Chapter 9).

**TABLE 17** Baseline risk of neonatal adverse events

Gestational age (weeks)	Adverse event					
	Death		RDS		IVH	
	Event probability	Probability distribution [Beta( $\alpha,\beta$ )]	Event probability	Probability distribution [Beta( $\alpha,\beta$ )]	Event probability	Probability distribution [Beta( $\alpha,\beta$ )]
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(687,43773)	0.03	Beta(1458,43002)	0.00	Beta(0,44460)

**Note**

Source: UK stillbirth and neonatal mortality rates (ONS 2016<sup>13</sup>) and US data from Medscape (<https://emedicine.medscape.com/article/260998-overview#a5>; accessed November 2017). The adjusted analysis used data on ACS treatment effects from Travers *et al.*<sup>93</sup> and steroid use in routine practice from the Neonatal National Audit Programme.<sup>94</sup>

**FIGURE 11** Baseline risks of neonatal adverse events. Source: UK stillbirth and neonatal mortality rates (ONS 2016<sup>13</sup>) and rates of RDS and IVH in US data from Medscape (<https://emedicine.medscape.com/article/260998-overview#a5>; accessed November 2017).

equal to the unadjusted risk when no women are treated with ACSs or when ACSs have no effect on the risk (i.e.  $RR_{ANS} = 1$ ) and increases with ACS use and the effectiveness of ACSs to a maximum of  $1/(1 - P_{ANS})$  multiplied by the unadjusted risk (i.e. when relative risk = 0). According to the National Neonatal Audit Programme (NNAP),<sup>94</sup> the most representative data source on ACS 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 ACS. Because NNAP does not produce data by gestational age, we assume such value applies to all gestational ages. We identified one study (by Travers *et al.*<sup>93</sup>) that reports treatment effects of ACSs by gestational age (range 23 to 34 weeks) for death before discharge and used that to derive  $RR_{ANS}$  (Grant<sup>95</sup>) for neonatal death in *Table 23*; we adopted the ACS treatment effects on severe ICH by gestational age from the same source to approximate the  $RR_{ANS}$  for IVH. We could not find estimates for ACS treatment effects on RDS by gestational age and, thus, assumed a constant value of  $RR_{ANS}$  for this outcome, from the source described in *Steroid treatment* (Travers *et al.*<sup>93</sup> report treatment effects estimates for bronchopulmonary dysplasia, but the effects were so imprecisely estimated that point estimates implied that the  $RR_{ANS}$  was  $> 1$  for all but three gestational ages).

Because 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<sup>24</sup> and multiply the incidence of RDS and IVH by 1 minus the probability of neonatal mortality among neonates with these events, which is assumed to be constant across gestational ages. 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.<sup>96</sup> We found no new data on the probability of neonatal death among RDS cases and, thus, used the value from the NICE guidelines,<sup>24</sup> which was obtained from US data for 2004 [Centers for Disease Control and Prevention (CDC) 2007<sup>97</sup>]. The values are 0.054 [beta(875,15393) in probabilistic sensitivity analyses] for RDS and 0.205 [beta(76,394)] for IVH.

### **Steroid treatment**

In the model, women who test positive are treated with ANSs. In the scenarios in which women present at level 1 or 2 maternity hospitals at  $< 28$  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 ACS administration from results reported by the Effective Perinatal Intensive Care in Europe (EPICE) study,<sup>33</sup> a prospective cohort study that collected data from 19 regions in 11 European countries in 2011 and 2012. This study was selected as the largest and most representative source of data on ACS 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 4594 infants born at gestational ages of 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 ACSs may be effective even when administered up to 3 hours before delivery, which was expected to reduce mortality relative to no ACS by 26%.<sup>33</sup> The authors reported that 77.9% of the 1111 women who received ACSs  $< 24$  hours before delivery received only one dose of ACS. Treatment effects on IVH were also derived from this source. For treatment effects on RDS, we used data from the Cochrane Database Systematic Review of RCTs of the effectiveness of ACSs relative to no treatment or placebo,<sup>34</sup> which was also the source of values for sensitivity analyses. The main findings from the Cochrane review are summarised in *Table 18*. Because 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.

The effectiveness of steroids depends on the time from ANS administration to delivery. *Table 19* displays the treatment effect model parameter values for the base-case and sensitivity analysis, which reflect the reduced effects of ACSs when given  $\leq 7$  days before birth.



**TABLE 18** Treatment effect of ACSs

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

Data from Roberts *et al.*<sup>34</sup>

**TABLE 19** Estimated effectiveness values of steroid treatment used in the model

Parameter	Base-case value	Scenario analyses	Probabilistic distribution for sensitivity analysis [log-normal (mean, SD)]	Source
Treatment effects: neonatal mortality RR (95% CI)				
ACS ≤ 7 days vs. no ACS	0.5 (0.4 to 0.6)	0.69 (0.59 to 0.81)	Log-normal (0.50, 0.093)	Base-case analysis: ACS 24 hours a day, 7 days a week, adjusted estimate from <i>Table 2</i> in Norman <i>et al.</i> <sup>33</sup> Scenario analysis: Cochrane review <sup>34</sup> (did not distinguish by timing of ACS)
ACS > 7 days vs. no ACS	1	0.7 (0.6 to 0.9)  0.69 (0.59 to 0.81)	Not varied: fixed at 1	Scenario analysis: Cochrane review <sup>34</sup> and Norman <i>et al.</i> <sup>33</sup>
Treatment effects: RDS RR (95% CI)				
ACS ≤ 7 days vs. no ACS	1	0.66 (0.56 to 0.77)	Log-normal (0.66, 0.079)	Base-case analysis: Cochrane review <sup>34</sup>
ACS > 7 days vs. no ACS	1	0.66 (0.56 to 0.77)	Not varied: fixed at 1	Scenario analysis: Cochrane review 2017 <sup>34</sup>
Treatment effects: IVH RR (95% CI)				
ACS ≤ 7 days vs. no ACS	0.6 (0.5 to 0.9)	0.55 (0.40 to 0.76)	Log-normal (0.60, 0.207)	Base-case analysis: ACS 24 hours a day, 7 days a week, adjusted estimate from <i>Table 2</i> in Norman <i>et al.</i> <sup>33</sup>  Scenario analysis: Cochrane review <sup>34</sup>
ACS > 7 days vs. no ACS	1	0.55 (0.40 to 0.76)  0.8 (0.6 to 1.2)	Not varied: fixed at 1	Scenario analysis: adjusted estimate from <i>Table 2</i> in Norman <i>et al.</i> , <sup>33</sup> and Cochrane review <sup>34</sup>
Treatment effects: birthweight mean difference (g) (95% CI)				
ACS ≤ 7 days vs. no ACS	0	0	Not varied: fixed at 0	Assumption based on Roberts and Dalziel <sup>32</sup>
ACS > 7 days vs. no ACS	0	-147.0 (-292.0 to -2.0)	Not varied: fixed at 0	Base-case analysis: assumption based on low quality of evidence in Roberts and Dalziel <sup>32</sup>  Scenario analysis: WHO <sup>98</sup> and Roberts and Dalziel <sup>32</sup>

RR, relative risk.

## Health-related quality of life

We conducted a systematic search of the literature for utility values of neonatal outcomes in the model: mortality, RDS and IVH. The details of the search strategy, identification and data extraction from the identified studies are provided in *Appendix 4*, in which additional tables highlight the parameter values used in the model. In this section, we summarise our findings.

### *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 4*.

Of these 28 studies, 24 assess the outcomes of children born preterm. The details of these studies are summarised in *Appendix 5, Table 39*. Nine additional papers were cited as sources for utilities in some of these studies. Parameter values from these additional papers are presented in *Appendix 5, Table 40*. The remaining four papers assess the outcomes of mothers; these studies are summarised in *Appendix 5, Table 41*.

### *Short Form questionnaire-36 items 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 that allow Short Form questionnaire-36 items (SF-36)<sup>99</sup> summary measures to be converted into EQ-5D utilities.<sup>100</sup> We made use of a mapping function obtained from Rowen *et al.*<sup>101</sup> to undertake 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 4*.

### *Relevant studies for utilities of intraventricular haemorrhage and respiratory distress syndrome 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.<sup>102</sup> However, a suitable mapping to EQ-5D utility for the health-related quality-of-life measure used in this paper could not be found.

Likewise, only one paper<sup>103</sup> considers quality of life for preterm children with RDS. This study measures SF-36 scores, but does not report them. We were unable to obtain the SF-36 data after contacting the corresponding author.

The evidence on the quality of life of mothers of preterm children is sparse. Only two studies consider mothers of preterm children specifically.<sup>104,105</sup> The first is an abstract that reported only physical and mental health SF-36 mean summary scores,<sup>104</sup> and the second reported Maternal Postpartum Quality of Life (MAPP-QOL) scores.<sup>105</sup> Neither of these could be reliably mapped to EQ-5D utilities.

Couto *et al.*<sup>106</sup> assessed the quality of life for mothers in Brazil who have had at least one of four previous adverse pregnancy outcomes. Although 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 4*.

### *Utilities for reduced birthweight*

In order to assess whether or not 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 data and utility data. One study was identified, and the author was able to provide the raw data on request.<sup>107</sup>

A number of regression specifications were estimated using random-effects estimators, in order to find the effects of birthweight on utility (see *Appendix 5, Statistical analysis of the effects of birthweight on utility*). 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 LR test when compared with a specification that included gestational age, sex and time dummies. Therefore, the analysis found insufficient evidence of a birthweight effect on utility, and so we assumed that there is no utility loss from reduced birthweight alone. Further details of the statistical analysis can be found in *Appendix 5, Statistical analysis of the effects of birthweight on utility*.

### Utility parameters selected for the economic model

Based on the discussion in *Appendix 4*, utilities deemed the most appropriate base-case values are reported in *Table 20*. Proxy utilities for RDS and IVH were obtained from the paper by Carroll and Downs,<sup>108</sup> which was the source for the study by Bastek *et al.*<sup>109</sup>

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 applied this proxy utility to 56% of all RDS cases. This figure is from a UK-based study that found that 56% of children born extremely preterm had abnormal baseline spirometry when they were 11 years old.<sup>110</sup> 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 of below III. On consultation with a clinical expert, long-term outcomes from IVH grades below III are thought to not differ greatly from those of preterm survivors in general. Therefore, we apply the utility of moderate cerebral palsy only to incidences of IVH that are at grade III or IV.

The utility for preterm survivors was obtained using mapped SF-36 scores.<sup>101</sup> The utility computed from the UK study<sup>111</sup> was used for the base case, and the minimum and maximum utilities in the remaining papers were selected to provide a range.<sup>112,113</sup> This follows the principles outlined in the NICE technical support document.<sup>114</sup>

**TABLE 20** Utilities selected for the economic model

Variable	Subject	Source	Measure	Utility	Range
'Severe' RDS (severe persistent asthma used as proxy)	Child	Carroll and Downs <sup>108</sup>	TTO	0.85	0.84–0.86 <sup>a</sup>
IVH grades III–IV (moderate cerebral palsy used as proxy)	Child	Carroll and Downs <sup>108</sup> via Bastek <i>et al.</i> <sup>109</sup>	TTO	0.76	0.66–0.84 <sup>b</sup>
Death	Child	Assumption. Upper bound from Vandenbussche <i>et al.</i> <sup>115</sup>	SG (upper bound only)	0	0–0.02
Preterm survivor	Child	Cooke <sup>111</sup>	SF-36 <sup>c</sup>	0.879	0.846–0.901 <sup>d</sup>
Mother with previous adverse child outcome	Mother	Couto <i>et al.</i> <sup>106</sup>	SF-36 <sup>c</sup>	0.644	0.556–0.652 <sup>e</sup>
Mother with no adverse child outcome	Mother	Couto <i>et al.</i> <sup>106</sup>	SF-36 <sup>c</sup>	0.834	0.768–0.843 <sup>e</sup>

SG, standard gamble; TTO, time trade-off.

a Range calculated as a 95% CI, based on the data from Carroll and Downs.<sup>108</sup>

b Range taken directly from the study by Bastek *et al.*<sup>109</sup> These represent the minimum and maximum values found in their literature search for these utilities.

c SF-36 means for the eight dimensions were mapped onto EQ-5D utilities using a quadratic model.<sup>101</sup>

d Range taken from two of the five studies reporting SF-36 scores (the minimum and maximum utilities reported among the five studies).<sup>112,113</sup>

e Range is generated by generating a 95% CI for the eight SF-36 means, and mapping all lower bounds and all upper bounds to EQ-5D using a linear model.<sup>101</sup> These represent a wider estimate of the range than if a 95% CI was provided for the EQ-5D measure directly.

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 the study by Couto *et al.*<sup>106</sup> This can be used as a proxy for mothers who have preterm children who suffer an adverse outcome, with the caveat that it is likely to be an overestimate for infant mortality and an underestimate for IVH or RDS. The lower (upper) bound for the two utilities imputed from the study by Couto *et al.*<sup>106</sup> is calculated by taking the lower (upper) value of the 95% CI 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.

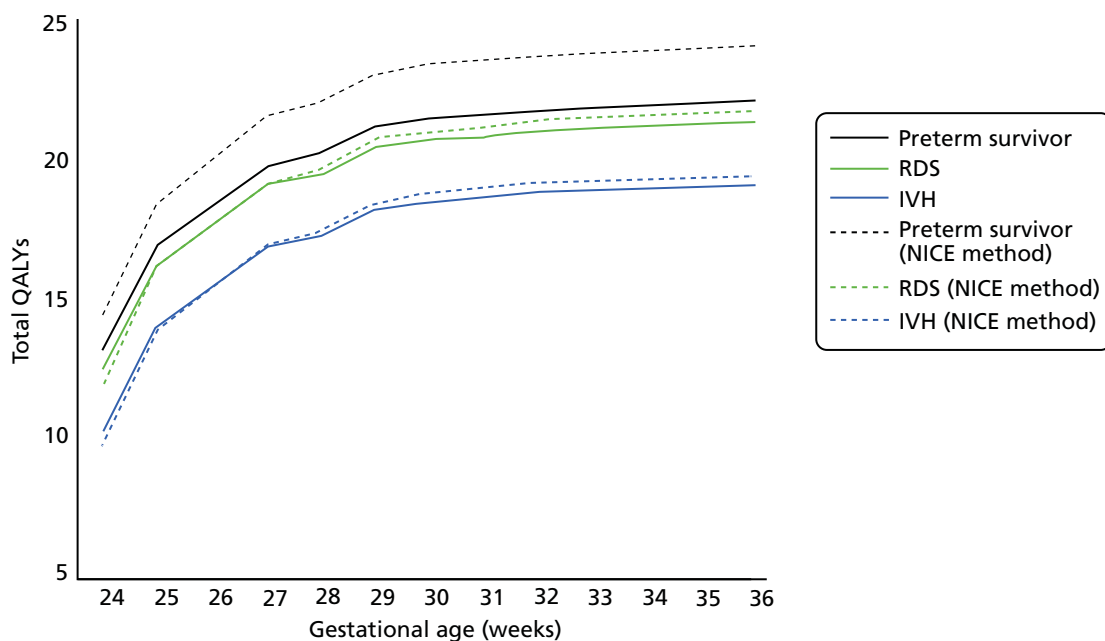
Therefore, the utilities of mothers and infants were linked by the occurrence of neonatal mortality and morbidity events both in deterministic and stochastic analyses. Further details on how discounted QALYs were obtained from these utility values, and a comparison of these to the QALY values used in the NICE guidelines model,<sup>24</sup> can be found in *Appendix 5, Quality-adjusted life-year calculations and comparison with parameters used in the National Institute for Health and Care Excellence guidelines model*. Details may be found therein for the basis of assumptions such as the 1-year duration of mortality effects.

### Summary of all child-related quality-adjusted life-year values used in the model

The base-case total discounted lifetime QALYs for children are summarised in *Figure 12*. For comparison, total discounted lifetime QALYs calculated using the NICE guidelines method<sup>24</sup> (i.e. without age and survival adjustments across the lifespan) are denoted by dashed lines in *Figure 12*. As can be seen by comparing these values, the NICE guidelines method for obtaining QALYs appears to:

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

This illustrates that the NICE guideline method is likely to overestimate the benefits of treatment-intensive options. It must be noted that this comparison of the two methods to extrapolating outcomes is undertaken within the Peninsula Technology Assessment Group (PenTAG) model structure, which is based on the effects of ACS treatment, as opposed to the original NICE model's<sup>24</sup> tocolytic-based mechanism



**FIGURE 12** Total lifetime QALYs for preterm children by gestational age and health state, using age and survival adjustment (solid lines) and NICE's unadjusted method<sup>24</sup> (dashed lines). 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.

of effect. Therefore, a comparison of the overestimation of QALY gains from treatment-intensive options (e.g. treat all vs. qfFN) in the original NICE model<sup>24</sup> is slightly smaller because, as discussed in *Chapter 5* the NICE model underestimated the benefits of treatment as it was based on tocolysis as opposed to ACSs.

## Costs

The cost parameter values used in the model are presented in *Table 21*. 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 ACS 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 2009/10 in a London university hospital.<sup>116</sup> These costs were inflated to 2016 prices using the Health Care and Prices Index in the Personal Social Services Research Unit's publication of *Unit Costs of Health and Social Care*.<sup>117</sup> The cost of in-utero transfer in this study only includes 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.

**TABLE 21** 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 Curtis and Burns <sup>117</sup>  Costs of the test excluding VAT (typically £35) (request for information from NICE to Hologic, Inc., 2017, personal communication from NICE). Owing to lack of data, this figure does not include the cost of test failures
plGFBP-1	35	Based on 10 minutes of midwife time, with unit costs from Curtis and Burns <sup>117</sup> and £15 cost of Medix test excluding VAT (request for information from NICE to Alere Inc., 2017, personal communication from Alere Inc.)
PAMG-1	52	Based on 10 minutes of midwife time, with unit costs from Curtis and Burns. <sup>117</sup> Cost of the test in the UK excluding VAT (request for information from NICE to Parsagen Diagnostics Inc., 2017, personal communication from Parsagen Diagnostics Inc.)
Maternal steroid injection	5	UCLH 2012 from Parisaei <i>et al.</i> <sup>116</sup>
Atosiban plus Atosiban infusion equipment	362	Atosiban infusion equipment includes syringe pump, syringe and giving set (Parisaei <i>et al.</i> <sup>116</sup> ). Dosage or units of doses not given, nor does it include costs of time to administer. Alternative values: BNF 2016, cost of solution for 37.5 mg/5 ml of infusion Atosiban acetate concentrate for solution for infusion vials, one vial costs £52.82 (hospital only) at maximum dose or alternatively half the maximum adult dose in BNF 70 of 330.75 mg over 48 hours plus equipment cost (Parisaei <i>et al.</i> <sup>116</sup> )
Inpatient hospital	1325	Median length of hospital stay (2 days) multiplied by the cost of 24-hour admission to hospital (Parisaei <i>et al.</i> <sup>116</sup> based on data from Primary Care Trust in London)
In-utero transfer	965	London Ambulance Service, 2012 (Parisaei <i>et al.</i> <sup>116</sup> ). It includes 6 hours (Gale <i>et al.</i> ) <sup>118</sup> of a modern matron's time to arrange transfer [i.e. 6 multiplied by £62 (Curtis and Burns <sup>117</sup> )]
Long-term health-care costs of IVH	114,648	Downstream health-care costs. NICE guideline 2015. <sup>24</sup> 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 <sup>24</sup> seems to be wrong; we assume that the correct number is equal to ≥ £79,000, and use this number when all grades of IVH are considered (alternatively, the assumption in the NICE 2015 guideline model <sup>24</sup> 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 <sup>119</sup> and for a discount rate of 3.5%

continued

**TABLE 21** Cost parameter values (2016 prices) (*continued*)

Cost parameter	Unit cost (£)	Definition, price-year and source
Neonatal hospital costs of preterm survivors discharged home/to ward	32,435	PenTAG analysis of Badger data <sup>120,121</sup> for infants born at gestational ages of < 36 weeks in England and Wales in 2013/14 ( $n = 22,936$ ). Includes the costs of BAPM levels 1–5 (XA01Z, XA02Z, XA03Z, XA04Z and XA05Z) at 2014/15 NHS tariffs. The mean overall length of stay (superspell) was 46 days (potential outcome without death)
Neonatal hospital costs of RDS	5587	OLS-adjusted difference in neonatal hospital costs between infants with and without days spent in BAPM level of care 1 in Badger 2014/15 data; <sup>120,121</sup> valued at the national tariffs for BAPM levels 1–5 (XA01Z, XA02Z, XA03Z, XA04Z and XA05Z) in 2014/15 prices. Alternative value: downstream health-care costs; <i>NHS Reference Costs 2011/12</i> , <sup>79</sup> XB01Z Paediatric Critical Care, Intensive Care, ECMO/ECLS; NICE guideline 2015. <sup>24</sup> Alternative value 2: Landry <i>et al.</i> , <sup>122</sup> preterm hospitalisation cost: 2008 Canadian dollars; cost of medical and pharmaceutical services also given. Adjusted using the HCHS indices and the purchasing power parities to reflect the equivalent costs in the UK in Great British pounds in 2016. The HCHS index for 1990/91 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. <sup>120,121</sup> Alternative value: assumption of no costs, as in NICE 2015 guideline model. <sup>24</sup> Alternative value 2: from Khan <i>et al.</i> <sup>123</sup>

HCHS, Hospital and Community Health Service; OLS, ordinary least squares; UCLH, University College London Hospitals.

**Note**

The costs were adjusted using the HCHS indices<sup>117</sup> to reflect the equivalent prices in 2016.

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 2014/15.<sup>124</sup> 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). 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.<sup>125</sup>

Unlike the economic analysis that informed the NICE 2015 guidelines<sup>24</sup> 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)<sup>46</sup> that saving a baby by means of timely ACS 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 the age of 11 years, it is estimated that neonatal hospital costs account for ≈90% of the total.<sup>126</sup>

The costs of RDS and IVH were estimated from the Badger data set.<sup>120,121</sup> In the data set 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 vs. induced) and mode of delivery (vaginal vs. caesarean section), 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 can also be for other causes, such as sepsis, necrotising enterocolitis and pulmonary haemorrhages, and may

be for reasons as varied as nutrition, surgery, chest drains and congenital abnormalities. However, some of these other causes may be affected by ACSs but are not amenable to be formally accounted for in our model owing to a 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 ACSs in our model.

### 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. 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 at 12.1%.<sup>92</sup> In order to derive the costs and QALYs of PartoSure for comparison with fFN at 50 ng/ml (and other options in the full incremental analyses), we conducted an indirect comparison whereby the incremental costs and QALYs of PartoSure relative to fFN at 50 ng/ml based on the test accuracy data from the study by Hadzi-Lega *et al.*<sup>44</sup> were added to the costs and QALYs of Actim Partus based on the APOSTEL-1 data.<sup>42,43</sup> *Table 22* summarises the main model assumptions.

### 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 that ACSs have (partial) benefits when administered earlier than 7 days before preterm delivery
- excluding the neonatal hospital costs of infant death
- including mothers' QALYs.

### Probabilistic sensitivity analyses

We present probabilistic analyses using information on sampling uncertainty for test accuracy, costs and utilities presented in *Tables 16, 17* and *19*. These analyses are presented in terms of the relative frequency with which each diagnostic test option had the highest net monetary benefit of all competing options.<sup>127</sup>

## Results

Because our results vary by length of gestation, we present details for the base case of a symptomatic woman presenting at 30 weeks' gestation (the average age on diagnostic accuracy studies) and general results for longer (33 weeks) and shorter (26 weeks) gestation.

### Base-case results

The base-case deterministic results are presented in *Table 23*. These are based on the preferred comparative studies APOSTEL-1<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup> The base case considers women presenting at 30 weeks' gestation to a level 2 hospital. Although all incremental cost-effectiveness ratios (ICERs) are positive, they should be interpreted with caution because, other than 'treat all' and fFN of 10 ng/ml versus fFN of 50 ng/ml, they represent a reduction in both costs and QALYs. Actim Partus results in £56,030 of cost savings per QALY lost relative to fFN of 50 ng/ml, which are higher than those of fFN of 200 ng/ml (£25,209) and fFN of 500 ng/ml (£17,025). The costs savings and QALY losses for PartoSure versus fFN of 50 ng/ml are the result of an indirect comparison between the studies by Bruijn *et al.*<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup> because 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 of 50 ng/ml.

TABLE 22 The PenTAG model specifications and assumptions

Model feature/parameter	Base-case model specification/assumption	Comment
Patient population	Symptomatic women with intact membranes presenting to a level 2 hospital, who have not been ruled for preterm labour after clinical examination	Clinical examination is not a relevant comparator for this population because 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 ACSs. Maternity costs are dependent on test sensitivity [i.e. false negatives (i.e. delivering within 7 days after a negative test result) are 'missed' (do not receive ACS)]	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 depend only on test specificity)
Adverse events included neonatal mortality and morbidity	Outcomes considered are neonatal death, RDS and IVH	This is in line with the previous model informing the NICE guidelines on preterm labour diagnosis and treatment (NICE 2015 <sup>24</sup> )
Neonatal mortality results in net savings to the NHS	Saving a neonatal life through accurate diagnosis and timely ACS treatment, has the consequence of increasing NHS costs, because the infant saved stays longer in neonatal hospital (although at a slightly less intensive average level of care per day)	In scenario analyses, we explore the impact on results of assuming that saving a child does not incur additional costs
ACSs are effective only if given within 7 days of delivery	Infants born more than 7 days after testing positive do not benefit from ACSs	In scenario analysis, we allow for partial benefit from ACSs for those testing positive and given ACS earlier than 7 days before preterm birth (i.e. at < 37 weeks)
In utero transfers are only required for women presenting to a level 1/2 hospital at < 28 weeks' gestation	Transfer to a tertiary hospital	This is in line with NICE guidelines <sup>24</sup>
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–risk profile of tocolysis
Long-term costs	Only included those associated with IVH	In line with the model informing the NICE 2015 guideline <sup>24</sup> of preterm labour diagnosis and treatment
Long-term quality of life	Assumed that those who survive beyond 1 year of life 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 <sup>24</sup> of preterm labour diagnosis and treatment; plausibility supported with clinical experts' opinion



**TABLE 23** Summary of ICERs for women presenting at 30 weeks' gestation (at a level 2 hospital)

Test	Total costs	Total QALYs	Base-case results					
			vs. treat all			vs. fFN at 50 ng/ml		
			Incremental costs	Incremental QALYs	ICER (per QALY)	Incremental costs	Incremental QALYs	ICER (per QALY)
Actim Partus <sup>a</sup>	£4891	22.010	–£1116	–0.010	£108,319 <sup>b</sup>	–£346	–0.006	£56,030 <sup>b</sup>
PartoSure <sup>c</sup>	£4731 <sup>d</sup>	22.010 <sup>c</sup>	–£1110	–0.008	£140,587 <sup>b</sup>	–£506	–0.006	£81,922 <sup>b</sup>
Treat all	£6007	22.020	£0	0	–	£770	0.004	£186,754
fFN <sup>a</sup>								
10 ng/ml	£5526	22.018	–£481	–0.002	£233,241 <sup>b</sup>	£289	0.002	£140,267
50 ng/ml	£5237	22.016	–£770	–0.004	£186,754 <sup>b</sup>	£0	0	–
200 ng/ml	£4995	22.006	–£1012	–0.014	£73,673 <sup>b</sup>	–£242	–0.010	£25,209 <sup>b</sup>
500 ng/ml	£4840	21.992	–£1167	–0.027	£42,485 <sup>b</sup>	–£398	–0.023	£17,025 <sup>b</sup>

a Bruijn *et al.*<sup>42,43</sup>

b The ICER represents the south-west quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

c Hadzi-Lega *et al.*<sup>44</sup> for comparison with treat all, indirect comparison between Bruijn *et al.*<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup> for comparison with fFN at 50 ng/ml (Bruijn *et al.*<sup>42,43</sup> was used as the reference study in this case).

d Inferred total cost and QALYs for PartoSure were obtained by applying relative differences vs. Actim Partus found using Hadzi-Lega *et al.*<sup>44</sup> to Bruijn *et al.*<sup>42,43</sup>

#### Note

Green text denotes ratios in which the difference in cost and the difference in benefit (QALYs) are both negative relative to the reference test option.

Table 24 shows the base-case results as a full incremental analysis. The rows of the table are ordered from the most to the least effective testing option in terms of total QALYs. Incremental costs, QALYs and cost-effectiveness for each test are shown in comparison with the following option in the table; for example, fFN at 50 ng/ml is immediately above Actim Partus in the QALY ranking, but has lower costs than this test and therefore dominates it. The relevant ICER of fFN at 50 ng/ml is, therefore, the ICER calculated relative to PartoSure (i.e. £81,922; see Table 23). A graphical depiction of these results is presented in Figure 13.

Table 25 breaks down the base-case results shown in Table 23 in terms of their component discounted costs and QALYs. It should be noted, as in Table 23, that the cost saving per QALY loss for PartoSure versus fFN at 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 the study by Hadzi-Lega *et al.*<sup>44</sup> were applied to the results for Actim Partus using the study by Bruijn *et al.*<sup>42,43</sup> and then compared with fFN at 50 ng/ml.

The ICERs for women presenting at 26 weeks' gestation and 33 weeks' gestation at a level 2 hospital are provided in Appendix 6.

### 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 versus the comparator of fFN at 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 *et al.*<sup>43</sup> (APOSTEL-1) study and the Hadzi-Lega *et al.*<sup>44</sup> study are presented here and in Appendix 6.

**TABLE 24** Fully incremental analysis of ICERs for women presenting at 30 weeks' gestation at a level 2 hospital

Test	Total costs	Total QALYs	Base-case results vs. next option in the QALY ranking		
			Incremental costs	Incremental QALYs	ICER
Treat all (test none)	£6007	22.020	£481	0.002	£233,241
fFN at 10 ng/ml <sup>a</sup>	£5526	22.018	£289	0.002	£140,267
fFN at 50 ng/ml <sup>a</sup>	£5237	22.016	£346	0.006	£81,922 <sup>b</sup>
Actim Partus <sup>a</sup>	£4891	22.010	£160	0.000	Dominated by PartoSure
PartoSure <sup>c</sup>	£4731 <sup>d</sup>	22.010 <sup>d</sup>	-£264	0.003	Dominates fFN at 200 ng/ml and fFN at 500 ng/ml
fFN at 200 ng/ml <sup>a</sup>	£4995	22.006	£155	0.014	£11,296 (dominated by PartoSure)
fFN at 500 ng/ml <sup>a</sup>	£4840	21.992	-	-	- (dominated by PartoSure)

a Bruijn *et al.*<sup>42,43</sup>

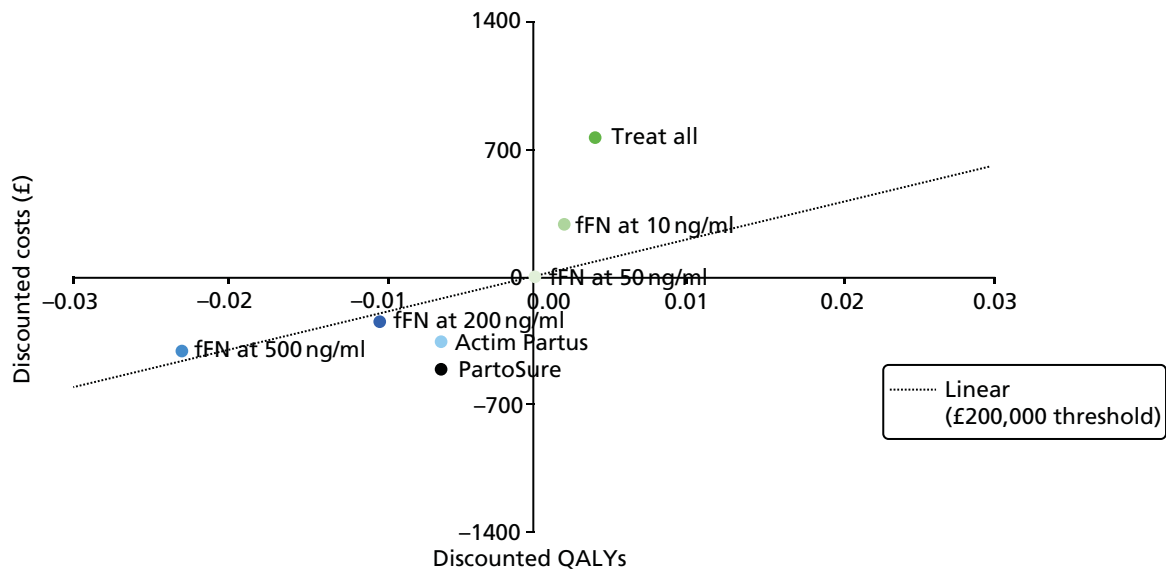
b Calculated relative to PartoSure, which dominates Actim Partus.

c Hadzi-Lega *et al.*<sup>44</sup> for comparison with treat all, indirect comparison between Bruijn *et al.*<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup> for comparison with fFN at 50 ng/ml (Bruijn *et al.*<sup>42,43</sup> was used as the reference study in this case).

d Inferred total cost and QALYs for PartoSure were obtained by applying relative differences vs. Actim Partus found using Hadzi-Lega *et al.*<sup>44</sup> to Bruijn *et al.*<sup>42,43</sup>

**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).



**FIGURE 13** Incremental costs and benefits of index tests against comparator (fFN at 50 ng/ml).

TABLE 25 Breakdown of base-case results (discounted costs and QALYs)

Cost and benefit components	Base-case results						Indirect comparison <sup>a</sup>	
	Treat all	Bruijn <i>et al.</i> <sup>42,43</sup> (APOSTEL-1)				Actim Partus		PartoSure
		fFN at 10 ng/ml	fFN at 50 ng/ml	fFN at 200 ng/ml	fFN at 500 ng/ml			
Discounted costs (£)								
Diagnosis	0	66	66	66	66	35	52	
Treatment	5	3	2	1	0	1	0	
Hospital admission	1325	781	493	250	95	177	1	
In-utero transfer	0	0	0	0	0	0	0	
Neonatal IVH	4006	4008	4010	4018	4030	4015	4015	
Neonatal RDS	624	624	625	627	630	626	626	
Neonatal death <sup>b</sup>	47	45	43	33	20	36	36	
Total	6007	5526	5237	4995	4840	4891	4731	
Incremental costs (vs. fFN at 50 ng/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	
Infant morbidity due to IVH	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	
Infant morbidity due to RDS	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	
Infant mortality avoidance	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 at 50 ng/ml)	0.004	0.002	Reference	-0.010	-0.023	-0.006	-0.006	
ICER vs. fFN at 50 ng/ml (£)	186,754	140,267	Reference	25,209 <sup>b</sup>	17,025 <sup>c</sup>	56,030 <sup>c</sup>	81,922 <sup>c</sup>	

a Costs and QALYs are inferred values computed via an indirect comparison between the Hadzi-Lega *et al.*<sup>44</sup> and Bruijn *et al.*<sup>42,43</sup> studies.

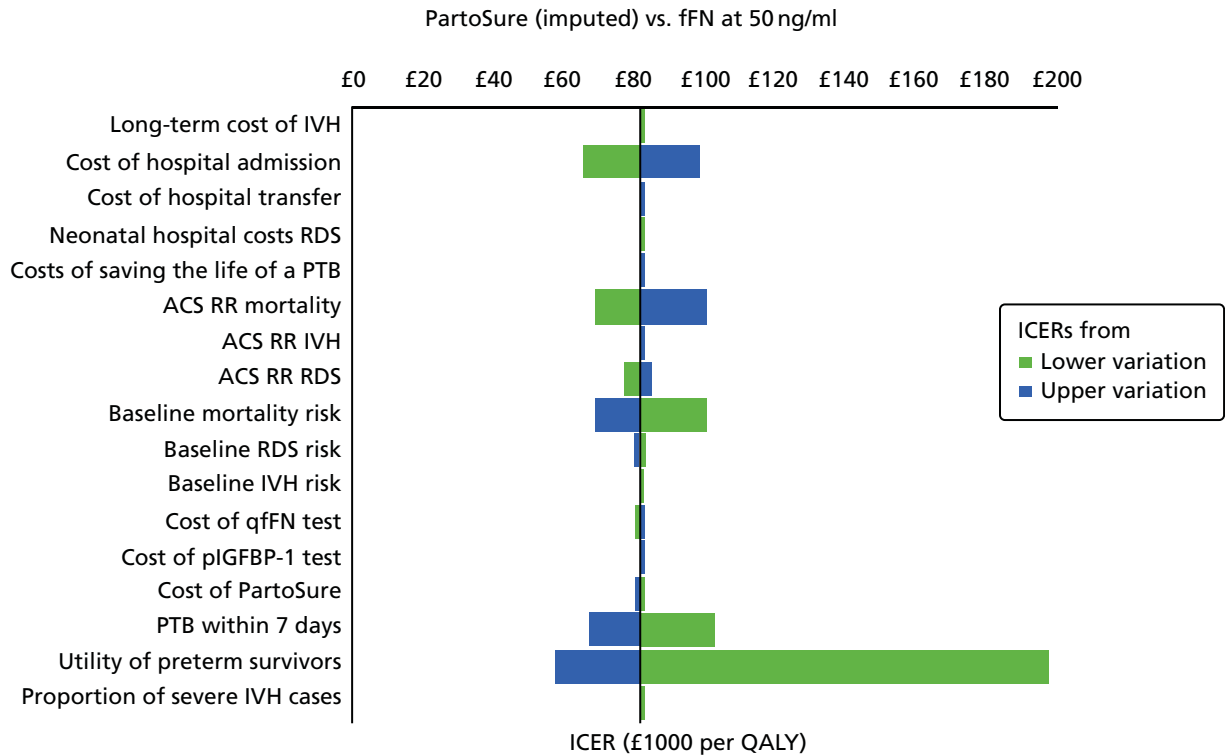
b These are the neonatal hospital costs associated with those infants saved by steroid treatment.

c The ICER represents the south-west quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

**Note**

Green text denotes ratios in which the difference in cost and the difference in benefit (QALYs) are both negative relative to the reference test option.

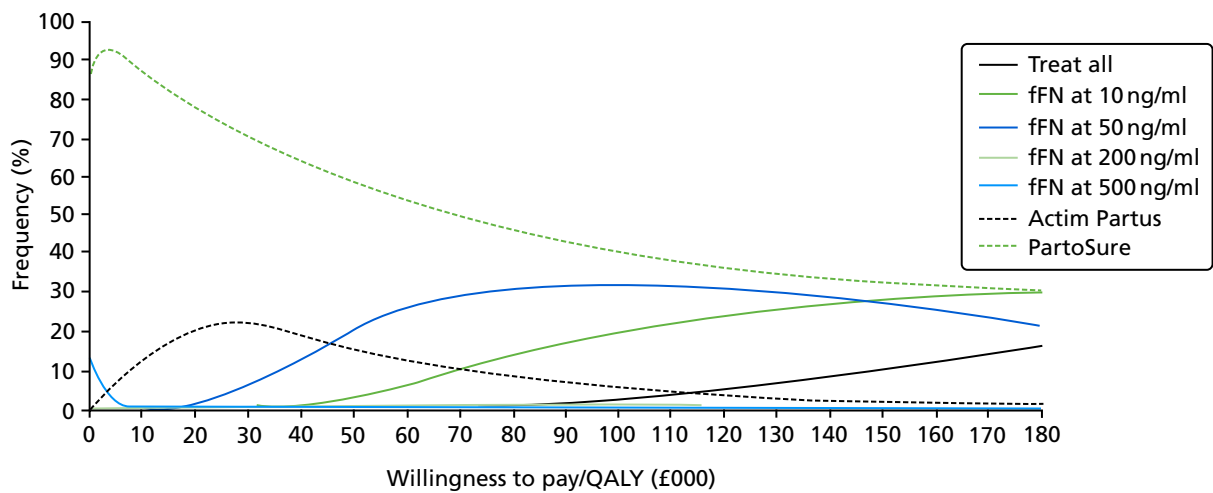
The tornado plot for PartoSure (imputed) versus fFN at 50 ng/ml is shown in *Figure 14*. The remainder can be found in *Appendix 6* (see *Figures 19–23*). There is a consistent pattern across all comparisons. The results are sensitive to the health-related quality of life (state utility) of preterm survivors. 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 14** Tornado diagram for model parameters with the most influential variation in value: PartoSure. PTB, preterm birth; RR, risk ratio.

**Probabilistic sensitivity analyses**

At a willingness to pay per QALY threshold of £20,000, Actim Partus has a probability of being cost-effective of 14%, 21% and 21% for women presenting at gestational ages of 33, 30 and 26 weeks, respectively (Figure 15, and see Appendix 6, Figures 24 and 25). PartoSure has probabilities of being cost-effective of 83%, 76% and 75% for women presenting at gestation of 33, 30 and 26 weeks, respectively, but these values are based on indirect comparison and have a lower strength of evidence than for other diagnostic options.



**FIGURE 15** Probabilistic analysis: women presenting at 30 weeks' gestation. The highest net monetary benefit is for a gestational age at presentation of 30 weeks.

The treat-all option has a probability of being cost-effective of 0% at a willingness to pay per QALY threshold of £20,000 for all women, and becomes the option with the highest likelihood of being cost-effective for women presenting at 26 weeks' gestation only at a willingness to pay per QALY threshold of > £180,000.

## Scenario analyses

### Alternative diagnostic accuracy data

Using the DTA results of the study by Cooper *et al.*,<sup>49</sup> which did not include PartoSure, suggests that the fFN at 50 ng/ml test provides lower costs with equal health benefit when compared with Actim Partus. The option of treating all women compared with fFN at 50 ng/ml yields an ICER of £34,508 per QALY.

(Confidential information has been removed.) (See *Appendix 6, Table 56.*)

When the diagnostic accuracy data from the meta-analysis of the four studies that compared Actim Partus with fFN at 50 ng/ml were used,<sup>42,43,51,56,57</sup> Actim Partus was dominant over fFN at 50 ng/ml as it resulted in cost savings of £41 per woman and health benefits of 0.01 more QALYs per woman. The treat-all option increased costs and QALYs and resulted in an ICER of £70,468.

### Other scenarios

Including the negative impact on QALY outcomes, the effect of an infant's death on mothers (which is assumed to last for 10 years) favours options that involve more use of ACS treatment. That is, in *Table 26* (and see *Appendix 6, Tables 63 and 64*), the ICER for treat all and fFN at 10 ng/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 delivery, the assessment becomes in effect a cost-minimisation analysis because our model does not account for health-related quality-of-life outcomes of the mother during the antenatal period. In this scenario, among women presenting at a gestational age of 30 weeks, PartoSure is the least costly option, with a £507 reduction in costs per woman, followed by fFN at 500 ng/ml (£400 reduction in costs per woman) and Actim Partus (£347 reduction in costs per woman). As discussed previously, the values for PartoSure need to be considered with caution.

When we allow for partial benefits of ACSs given earlier than 7 days before birth, the ICERs for fFN at 10 ng/ml and treat all are £24,420 and £41,625, respectively, among women presenting at 30 weeks' gestation; as for the rest, only PartoSure results in savings per QALY lost of > £20,000 relative to fFN at 50 ng/ml (see *Table 26*). Further details on the analyses of women presenting at a gestational age of 26 weeks are presented in *Appendix 6, Tables 54–56*. Detailed results for women presenting at 33 weeks' gestation are presented in *Appendix 6, Tables 57–59*. Tornado analyses of these gestational age groups are presented in *Appendix 6, Additional deterministic sensitivity analyses*, their associated probabilistic sensitivity analyses are presented in *Appendix 6, Additional probabilistic sensitivity analyses*, and other scenarios are presented in *Appendix 6, Scenario analyses*.

**TABLE 26** Incremental cost-effectiveness ratios vs. fFN at 50 ng/ml

Option	ICERs by scenario (£)									
	Base case	Presentation at 26 weeks' gestation	Presentation at 33 weeks' gestation	With maternal QALYs for 10 years	Limiting the analysis to delivery (additional cost only)	Limiting the analysis to the first year after birth	ACS earlier than 7 days before preterm delivery has partial benefits	Excluding additional neonatal hospital costs of death	Women presenting at a level 3 hospital	Applying costs and disutilities of AEs to all AEs
Treat all	186,754	128,939	323,093	111,813	770	4,930,356	41,625	185,771	186,754	175,158
fFN										
10 ng/ml <sup>a</sup>	140,267	92,845	242,716	74,564	289	3,704,141	24,420	139,284	140,267	131,558
200 ng/ml <sup>a</sup>	25,209 <sup>b</sup>	16,541 <sup>b</sup>	43,781 <sup>b</sup>	18,968 <sup>b</sup>	-243	669,219 <sup>b</sup>	9728 <sup>b</sup>	24,226 <sup>b</sup>	25,209 <sup>b</sup>	23,664 <sup>b</sup>
500 ng/ml <sup>a</sup>	17,025 <sup>b</sup>	11,476 <sup>b</sup>	29,631 <sup>b</sup>	13,347 <sup>b</sup>	-400	453,340 <sup>b</sup>	7428 <sup>b</sup>	16,042 <sup>b</sup>	17,025 <sup>b</sup>	15,968 <sup>b</sup>
Actim Partus <sup>a</sup>	56,030 <sup>b</sup>	35,364 <sup>b</sup>	97,069 <sup>b</sup>	38,200 <sup>b</sup>	-347	1,482,175 <sup>b</sup>	16,662 <sup>b</sup>	55,046 <sup>b</sup>	56,030 <sup>b</sup>	52,551 <sup>b</sup>
PartoSure <sup>c</sup>	81,922 <sup>b</sup>	53,446 <sup>b</sup>	141,838 <sup>b</sup>	81,893 <sup>b</sup>	-507	2,165,156 <sup>b</sup>	128,506 <sup>b</sup>	80,939 <sup>b</sup>	81,922 <sup>b</sup>	76,836 <sup>b</sup>

AE, adverse event.

a Bruijn *et al.*<sup>42,43</sup>

b The ICER represents the south-west quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

c Indirect comparison between the studies by Bruijn *et al.*<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup> (Bruijn *et al.*<sup>42,43</sup> was used as the reference study in this case).

**Note**

Green text denotes ratios in which the difference in cost and the difference in benefit (QALYs) are both negative relative to the reference test option.

## Chapter 7 Assessment of factors relevant to the NHS and other parties

- The practical considerations when conducting these tests, highlighted to us by our advising obstetricians, were:
  - The qfFN 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 cervical os as advised.
- Some of these tests have a dual purpose because they can be used for other indications (e.g. women with multifetal pregnancies, women with ruptured membranes and women not presenting with symptoms). Our population is not representative of the whole population of women presenting with threatened preterm labour. However, it is recognised that treatment of these other populations may involve different management strategies.
- It is highlighted that new diagnostic test options for diagnosing preterm labour and predicting preterm birth are likely to have spillover and feedback effects between maternity and neonatal care services. A proper evaluation of the potential impact of these tests would therefore need to account for the possibilities offered in the context of more integrated networks of maternity and neonatal services.
- There are potential implications of adopting new biochemical tests on neonatal unit workload and service planning.
- Health-care service and travel needs and costs are borne by patients and relatives owing 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 an inadequate level of specialisation, and the cost and health consequences associated with postnatal transfers.
- There are 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.
- There are long-term parental and societal economic impacts.
- There are equity implications of changes in hospital admissions across rural and urban areas.





## Chapter 8 Patient and public involvement

Patient and public involvement was not conducted for this project. This research project was part of the NICE Diagnostics Assessment programme in which, from the publication of the scope in week 12, there were 24 weeks in which to prepare the entire report. Regrettably, it was not logistical nor feasible to recruit and prepare for and report patient and public involvement in this project within the permitted time frame. Patients and the public were involved in the NICE Diagnostic Advisory Committee meetings, in which recommendations based on this report were made.



## Chapter 9 Discussion

### Review of test accuracy evidence

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

A summary of the key findings from the systematic review of test accuracy evidence can be found in *Chapter 2, Summary*. In brief, 20 included studies evaluated an index test against the 7-day reference standard<sup>41–62</sup> and seven studies evaluated an index test against a 48-hour reference standard.<sup>41,48,52,53,56–58</sup>

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

The studies that 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<sup>42,43</sup> (2016; assessing Actim Partus and qfFN) and the study by Hadzi-Lega *et al.*<sup>44</sup> (2017; assessing Actim Partus and PartoSure). No studies assessing qfFN and PartoSure within the same population were identified by our review. From APOSTEL-1, the sensitivity was superior for qfFN at the thresholds of 10 and 50 ng/ml compared with Actim Partus, whereas Actim Partus had a superior sensitivity compared with qfFN at the 200-ng/ml and 500-ng/ml thresholds. Specificity was superior using Actim Partus compared with qfFN at the thresholds of 10, 50 and 200 ng/ml but not against the threshold of 500 ng/ml. From the study by Hadzi-Lega *et al.*,<sup>44</sup> the sensitivities were the same for PartoSure and Actim Partus, whereas the specificity was superior using PartoSure. However, the CIs for all sensitivity and specificity data from the two tests assessed within these two studies overlapped considerably (all data available in *Table 9*).

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 (16 studies), the study with the best overall sensitivity and specificity results was that by Tripathi *et al.*<sup>57</sup> [sensitivity 94.7% (95% CI 89.9% to 97.7%) and specificity 92.4% (95% CI 88.0% to 95.1%)], whereas Cooper *et al.*<sup>49</sup> reported the worst sensitivity [33.3% (95% CI 4.3% to 77.7%)] and specificity [74.1% (95% CI 69.1% to 78.6%)]. These are two of the three largest studies identified in our review (Tripathi *et al.*,<sup>57</sup>  $n = 468$ ; Cooper *et al.*,<sup>49</sup>  $n = 349$ ). For PartoSure against the 7-day reference standard (four studies), the study with the best overall sensitivity and specificity results was that by Bolotskikh and Borisova<sup>59</sup> (2017) [sensitivity 100.0% (95% CI 73.5% to 100.0%) and specificity 95.4% (95% CI 88.6% to 98.7%)], whereas Werlen *et al.*<sup>41</sup> reported the worst sensitivity [0.0% (95% CI 0.0% to 97.5%)] and specificity [97.5% (95% CI 96.8% to 99.9%)]. The low sensitivity from the study by Werlen *et al.*<sup>41</sup> is attributable to only one woman delivering preterm (within 7 days) and her testing (falsely) negative within the study sample of size 41. If discounting the results from the study by Werlen *et al.*,<sup>41</sup> the next study reporting the worst overall sensitivity and specificity results was that by Nikolova *et al.*<sup>60,61</sup> (2015) [sensitivity 80.0% (95% CI 63.1% to 91.6%) and specificity 94.6% (95% CI 90.1% to 97.5%)]. The fFN at a threshold of 10 ng/ml (two studies) had a sensitivity range of 93.8% (95% CI 82.8% to 98.7%) to 95.7% (95% CI 87.8% to 99.1%) and a specificity range of 32.2% (95% CI 27.7% to 37.0%) to 42.3% (95% CI 36.5% to 48.4%). At a threshold of 200 ng/ml, fFN had a sensitivity range of 70.8% (95% CI 55.9% to 83.0%) to 71.0% (95% CI 58.8% to 81.3%) and a specificity range of 78.6% (95% CI 74.3% to 82.5%) to 83.6% (95% CI 78.8% to 87.8%). At a threshold of 500 ng/ml, fFN had a

sensitivity range of 29.2% (95% CI 17.0% to 44.1%) to 42.0% (95% CI 30.2% to 54.5%) and a specificity range of 94.3% (95% CI 91.6% to 96.4%) to 95.7% (95% CI 92.7% to 97.8%). Looking at these data, given the large ranges between studies assessing the same test and the wide CIs, it would be premature to attempt to deduce which test was superior to the 7-day reference standard.

We were able to assess Actim Partus and PartoSure only against the 48-hour reference standard because no studies were identified for qfFN. From the single PartoSure study (Werlen *et al.*<sup>41</sup>), the sensitivity was 0.0% (95% CI 0.0% to 97.5%) and specificity was 97.5% (95% CI 86.8% to 99.9%); the total sample size was 41 and only one test result was positive (a false positive). 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 48-hour reference standard (six studies), the study with the best overall sensitivity and specificity results was that by Tripathi *et al.*<sup>57</sup> [sensitivity 95.5% (95% CI 89.7% to 98.5%) and specificity 82.1% (95% CI 77.8% to 86.0%)], whereas Goyal *et al.*<sup>52</sup> reported the worst sensitivity [65.7% (95% CI 47.8% to 80.9%)] and specificity [56.0% (95% CI 34.9% to 75.6%)]. Given that we identified only 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 be premature to attempt to deduce which test was superior to the 48-hour reference standard.

We identified two relatively recent systematic reviews<sup>128,129</sup> that assessed fFN at a threshold of 50 ng/ml (i.e. current practice). Both of these reviews suffered from similar heterogeneity issues as our review and their summary ROC plots displayed large 95% prediction regions. The pooled sensitivities for the reviews by Boots *et al.*<sup>129</sup> and Sanchez-Ramos *et al.*<sup>128</sup> were 75% (95% CI 69% to 80%) and 76.1% (95% CI 69.1% to 81.9%), respectively, and the specificities were 79% (95% CI 76% to 83%) and 81.9% (95% CI 78.9% to 84.5%), respectively.

Both the study's heterogeneity and the uncertainty about the true accuracy of the index tests inevitably compromise the report's ability to firmly conclude whether or not the accuracy of the index tests is better than current practice.

### 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 prespecified 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 conducted by one reviewer and checked by a second.

### 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 dilatation threshold), inclusion of multiple gestations, participant characteristics and provision of treatments. As a consequence, the reported accuracies of individual tests varied widely, and hence the CIs 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 in which 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<sup>42,43</sup> (2016) and Hadzi-Lega *et al.*<sup>44</sup> (2017), assessed the DTA of two different index tests in the same population; APOSTEL-1<sup>42,43</sup> assessed both Actim Partus and qfFN and

Hadzi-Lega *et al.*<sup>44</sup> assessed Actim Partus and PartoSure. We did not identify any studies in which all three tests were used in the same population.

Our review was also limited by the lower number of published studies for qfFN (two studies) and PartoSure (four studies) than for Actim Partus (16 studies) against the 7-day reference standard. In addition, fewer studies published data against the 48-hour reference standard, with only seven studies being identified (six for Actim Partus and one for PartoSure). Meta-analysis of test accuracy data requires a minimum of four studies; therefore, the scope for meta-analysis was restricted. We are aware of three studies published after our searches were run that assessed PartoSure (Wing *et al.*,<sup>130</sup> Lofti *et al.*<sup>131</sup> and Melchor *et al.*<sup>132</sup>) and (confidential information has been removed). We also identified seven relevant ongoing trials, four of which are UK based, of which two plan to enrol > 1000 participants. Personal communications with trial organisers (Dr Sarah Stock, University of Edinburgh, 2017, and Professor Andrew Shennan, King's College London, 2017) indicated that data from the two large UK trials (QUIDS/QUIDS-2 assessing PartoSure, Actim Partus and qfFN and PETRA assessing qfFN) were expected in 2018. There is the potential, should our analyses be re-run using the data from these trials, that the estimates of relative test accuracy may change.

The scope issued by NICE<sup>12</sup> asked for an assessment of test accuracy of qfFN at thresholds other than 50 ng/ml. Our capabilities to look at different thresholds was limited by those reported in the published studies. The two qfFN studies both used thresholds of 10, 50, 200 and 500 ng/ml. Without access to the individual patient data, we were unable to assess any other thresholds.

Owing 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 qfFN studies. The second was to include studies in which 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 qfFN study (Bruijn *et al.*<sup>62</sup>).

Dependent on how the data were reported in each study, we were required to conduct some data manipulations. Most studies reported the raw true-positive, true-negative, false-positive and false-negative data, enabling us to calculate additional test accuracy statistics, such as sensitivity, specificity, PPV and NPV. However, four studies<sup>48,54,56,57</sup> reported only sensitivity, specificity, PPV and NPV. We back-calculated from these statistics to derive the true-positive, true-negative, false-positive and false-negative values, which were required in our review. All the raw data as reported in the published studies are available in *Appendix 3* (see *Tables 33* and *34*).

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 but that there may have been an impact on precision.<sup>133</sup> Our review did include a French study (Werlen *et al.*<sup>41</sup>), as a certified translation was received from the manufacturer (Parsagen Diagnostics Inc.).

### Areas of uncertainty

There is considerable uncertainty surrounding the generalisability of the studies to the UK population. Most specifically, none of the included studies was conducted in the UK. In addition, the prevalence rates of preterm birth in our included studies ranged from 2% to 73%. UK prevalence is approximately 8%. These differences in prevalence between studies are probably attributable 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 have an impact on the DTA data presented in *Chapter 2, Results of quantitative data synthesis (test accuracy data)*.<sup>134</sup> Indeed, Leeflang *et al.*<sup>134</sup> 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.

Our ongoing trial searches identified four relevant ongoing UK trials, two of which are very large (> 1000 participants) and their results are anticipated to be published in 2018 (PETRA and QUIDS).

There is some uncertainty around whether or not 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 on) cervical length measurement and two studies only included women with a transvaginal cervical length measurement of  $\leq 30$  mm. It is in these final two studies (APOSTEL-1<sup>42,43</sup> and Danti *et al.*<sup>50</sup>) in which there is most uncertainty; selection based on cervical length measurement would probably increase the prevalence of preterm birth in these studies. However, given that 15 of the 20 included studies had a prevalence rate of < 25% and both studies had a prevalence of < 25% (APOSTEL-1,<sup>42,43</sup> 19.7%; Danti *et al.*,<sup>50</sup> 6.7%), it is unlikely that these criteria had an impact on prevalence rates.

There was also uncertainty around whether or not any management strategies (e.g. treatments) would incorrectly inflate false-positive rates. As described in *Chapter 2, Summary of any treatments given*, 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 at 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 or not the mother is symptomatic, number of fetuses, gestation in weeks and days, qfFN value and/or cervical length measurement. In our review, we did not consider combining test results with such clinical data, because no studies were identified that assessed this combination.

We also acknowledge four studies of PartoSure and quantitative/qualitative fFN that have been recently published or are due to be published. These studies were published after our searches were run and consequently were not eligible for the review. The studies are:

- Wing *et al.*<sup>130</sup> (2017),  $n = 796$  – PartoSure and fFN at 50 ng/ml
- Lofti *et al.*<sup>131</sup> (2017),  $n = 132$  – PartoSure
- Melchor *et al.*<sup>132</sup> (2017),  $n = 420$  – PartoSure and fFN at 50 ng/ml
- (confidential information has been removed).

## Review of clinical effectiveness evidence

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

### 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 prespecified prospective protocol.

### 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 that have not been identified.

### Areas of uncertainty

Because no studies were identified for inclusion, we were unable to assess whether or not using these biomarker tests for predicting preterm labour is clinically effective (i.e. whether or not they would improve health outcomes).

### Review of cost-effectiveness evidence

Only one conference abstract was identified that was relevant to our cost-effectiveness review.<sup>66</sup>

### 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<sup>24</sup> 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.

Furthermore, 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<sup>24</sup> on preterm labour diagnosis and treatment, the evidence on the risk–benefit profile for tocolysis and steroids has changed. Tocolysis is now used sparingly, whereas the importance of providing steroids within 2 days of preterm delivery has gained consensus, especially following articles published in 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 at < 30 weeks' gestation should be admitted to hospital without testing, current practice has not followed this recommendation. Instead, the fFN and Actim Partus tests are commonly used to guide admission and treatment decisions in women as 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.

### Weaknesses

We focused on economic studies in symptomatic women with intact membranes, and did not cover studies of asymptomatic women, which might have provided important relevant evidence on utilities, costs, epidemiological parameters and modelling methods. 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 (dOes Progesterone Prophylaxis To prevent preterm labour IMprove oUtcoMe) trial by Norman *et al.*<sup>135</sup> in 2016]. We contacted the authors to request utility data, but 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 addressed only 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 qfFN in 1500 individuals across the UK (Dr Sarah Stock, University of Edinburgh, 2017, personal communication). This will constitute the largest known economic study to date on a biochemical test in the population of interest to this review, and will enable the investigation of key outcomes for clinical effectiveness, such as the ability to predict delivery within 2 days of presentation with symptoms suggestive of preterm labour. Previous diagnostic studies of fFN tests have provided limited data on such outcomes.

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 in which patients were not in preterm labour or likely to deliver within 7 days. More recent studies have extended the analysis from a purely cost-minimisation framework to a cost-effectiveness framework, by recognising that, however few false-negative 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<sup>24</sup> 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<sup>24</sup> on the topic, new evidence has emerged on the value of timely use of ACSs, which may still confer benefits on the neonate if given within 1 day of delivery and have a maximum benefit when given between 1 and 2 days of delivery. However, existing models are ill suited to account for this emerging evidence as few diagnostic studies are large enough to include sufficient numbers to measure such outcomes reliably. Current clinical guidelines and recommendations maintain that steroids need to be given within 7 days of delivery to be effective, although empirical findings published in 2017 by separate independent groups of researchers suggest residual beneficial effects when steroids are given more than 7 days before delivery.<sup>34,35,94</sup> On the other hand, existing studies have not accounted for the risks posed to the neonate by ACSs in terms of birthweight, despite the common perception by health professionals of the importance of these risks as manifested through the existing tendency to not give repeated courses of ACSs to women who do not deliver within 7 days of an initial course.

Finally, although the screening was carried out by two reviewers, the data extraction and critical appraisal were conducted by only one. This may increase the risk of bias associated with the findings of this review.

### **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 (a MSc dissertation<sup>68</sup>), 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 woman with steroids, as reported by the unpublished study of fFN, using a decision tree model and costs and utilities as used by the NICE 2015 guideline model.<sup>24</sup> In view of the limitations of the unpublished model highlighted previously, it is unclear whether or not these findings are robust to sampling and structural model uncertainty.

Key areas of structural uncertainty include the maintained assumption in the NICE model<sup>24</sup> 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 of experiencing adverse chronic events including RDS and IVH. In fact, some of those 'missed' cases are likely to return to the maternity hospital and receive ACSs 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, because previous studies have ignored the costs associated with neonatal deaths and have used low-quality data on costs of neonatal morbidity. Better quality data on neonatal hospital costs (length of stay at different levels of care) are available for the UK from the National Neonatal Research Database.<sup>124</sup>



## Independent economic assessment

In order to address some of the key limitations in the evidence base, primarily the lack of any evidence on the cost-effectiveness of the index tests in question, we developed a de novo model. The model incorporates the main elements of existing published models, in which 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 ACS 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,<sup>24</sup> 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 and qfFN at thresholds other than 50 ng/ml) with the comparators (fFN at 50 ng/ml and no testing and treating all) based on the best available diagnostic accuracy evidence for the tests. This turned out to be from the two available studies that compared at least two index tests in the same patient sample (Bruijn *et al.*<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup>). One study compared qfFN with Actim Partus in a group of women from the Netherlands ( $n = 350$ ),<sup>42,43</sup> and the other compared Actim Partus with PartoSure in a group of women from Macedonia ( $n = 37$ ).<sup>44</sup> Thus, we presented an economic evaluation of those two comparisons and also of the indirect comparison of PartoSure with fFN via Actim Partus as the common treatment to both studies, and included the no-testing treat-all and fFN at 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 at the expense of inferior health outcomes (fewer QALYs) relative to fFN at 50 ng/ml. 'No testing treat all' and fFN at 10 ng/ml had ICERs of  $> \text{£}100,000$  relative to fFN at 50 ng/ml, and fFN at 200 ng/ml and fFN at 500 ng/ml each produced less cost savings per unit of QALY lost relative to fFN at 50 ng/ml than Actim Partus. Probabilistic sensitivity analysis shows that the willingness to pay per QALY would need to be  $> \text{£}100,000$  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 the only UK study ( $n = 299$ ; Abbott *et al.*<sup>92</sup>) reported a 7-day prevalence rate of 3%, compared with 19.7% ( $n = 350$ ; APOSTEL-1, Bruijn *et al.*<sup>42,43</sup>) and 10.5% ( $n = 57$ ; Hadzi-Lega *et al.*<sup>44</sup>) 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 at 50 ng/ml than qfFN at 200 ng/ml and 500 ng/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.*<sup>49</sup>); however, this appears to have used a qualitative fFN test technology that is no longer in use.

In contrast to the results reported by the economic analysis that informed the 2015 NICE guidelines on preterm labour diagnosis and treatment,<sup>24</sup> we found that the policy of not testing and treating all women presenting at  $\leq 30$  weeks of gestation had ICERs well above the  $\text{£}20,000$ -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 ACSs. NICE assumed that the benefits were accrued 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. There were also differences in terms of the measured costs, because we included the costs to the NHS generated by infants whose lives were saved by ACS use, 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 risk values derived from those data for the fact that ACSs are now highly prevalent, with 83% of preterm infants born in the UK being given at least one ACS dose; the model used by NICE did not adjust for such prevalent use of ACS and, thus, may have underestimated the QALY benefits from treatment-intensive options.

### 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 ACS administration in terms of neonatal mortality and morbidity (RDS and IVH). Furthermore, we adopted the best modelling practice in the field and introduced some innovations by accounting for:

- gestational age
- hospital setting (level 1, 2 or 3)
- costs of in utero transfers at very preterm gestation
- 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.

### Weaknesses

We were unable to consider multipurpose 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 combinations of tests. The population was defined as women for whom a transvaginal ultrasound scan was not indicated or who were attending maternity hospitals in which it was unavailable.

We had to approximate the neonatal hospital costs of RDS and IVH, assuming that 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<sup>24</sup>).

Critically, the evidence on accuracy was limited to two comparative studies, one of which may be considered too 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 of Actim Partus and PartoSure. On firmer ground are the findings that, relative to fFN at 50 ng/ml, Actim Partus produced larger savings per QALY lost than fFN at 200 ng/ml and fFN at 500 ng/ml, whereas 'treat all' and fFN at 10 ng/ml had ICERs above the £80,000 mark in women presenting at 24–34 weeks' gestation; these results are based on diagnostic accuracy data from one study<sup>42,43</sup> in 350 women. However, it must be noted that our results were based on the assumption that the overall test accuracy reported in these studies is the same across gestational ages, owing to a lack of data by length of gestation.

It is worth noting that the only UK diagnostic study was a non-comparative study of qfFN, in which fFN at 200 ng/ml and at 500 ng/ml resulted in the same sensitivity as that of fFN at 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 to 200 ng/ml has an ICER of < £20,000, whereas lowering it from 200 to 50 ng/ml has an ICER of > £20,000.

There is an absolute lack of evidence on the outcomes of the status quo diagnostic test, qfFN, 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 6 months

(2016–17) 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 benefiting from being ‘missed’ because they may deliver between 2 and 7 days and, therefore, as an unintended consequence get the chance to return before delivery and closer to the target optimal window for ACS administration of 1 to 2 days before delivery.<sup>33</sup> Thus, many if not most false negatives may possibly end up having the same clinical outcomes as true positives. 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. because 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 previously. Another argument in support of cost minimisation is provided by the perceived harm of multiple courses of ACSs; in this case, the key diagnostic outcome measure for clinical effectiveness would be the FPR because not only would it result in ineffective use of ACSs to start with but it would measure the extent to which women were precluded from any use of ACSs and thus exposed to the risk of adverse neonatal health outcomes. This suggests that the cost-minimisation analysis may be thought of as a conservative alternative scenario to the base case.

We found no reliable data to account for the side effects of ACSs in terms of birthweight. The only source of data on long-term quality-of-life outcomes by birthweight that we identified was from a Canadian study by Saigal *et al.*,<sup>107</sup> which involved measured quality-of-life outcomes in 286 extremely low-birthweight survivors at adolescence and young and mature adulthood. Our analysis showed 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 at birth.

There was little evidence on quality-of-life outcomes of mothers, and we could conduct only some exploratory analysis of results including such evidence. Another limitation is that costs and utility values were independent of gestational age at birth; for example, we assumed that, provided a preterm infant survives until their first birthday, they will experience the same long-term quality of life as in the average population. Owing to a lack of data on the quality-of-life effects of IVH, we had to use utility values from other patient populations (i.e. individuals with ICH).

The mapping algorithm that we used to derive EQ-5D utility values from SF-36 produced predicted EQ-5D utilities that were consistently higher than (overpredicted) actual EQ-5D utilities for more severe EQ-5D states;<sup>101</sup> however, this prediction error is large only for health-state utilities of  $< 0.25$ . The lowest mapped utility we consider is 0.64, which was for mothers, and not used in base-case analysis. The only mapped utilities included in the base case were for preterm survivors, and these were all  $> 0.8$ . Therefore, although there is certainly a limitation with the mapping function, we do not believe that the limitation is a concern given the ranges of utility values we deal with in this review.

### 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 had access to only one comparative study of this test, which involved  $< 100$  individuals. A study of this size is unlikely to provide reliable results on diagnostic accuracy on delivery outcomes at 48 hours and 7 days. The 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 a common comparator (i.e. Actim Partus) approach.

More generally, as discussed previously, the ability to predict delivery within 48 hours has become critical, given recent evidence on the importance of good neonatal outcomes of administering ACSs 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 owing to a lack of the required data for the great majority of studies. Further DTA 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 the cost-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 ACSs may be accounted for in economic assessments.

There is a high degree of heterogeneity in the evidence on DTA, as illustrated by the very different prevalence of preterm birth between the largest comparative study (APOSTEL-1,<sup>42,43</sup> 20% delivery within 7 days) and the UK study (Abbott *et al.*,<sup>92</sup> 3% delivery within 7 days), which is not explained by the former study's inclusion and the latter's exclusion of non-spontaneous preterm births. More obvious differences in the selection of patients for study participation is also a problem, as highlighted by the high preterm birth prevalence in one of the four available studies comparing Actim Partus and fFN at 50 ng/ml,<sup>57</sup> which, by including patients at gestational ages of 28–36 weeks, rather than the more common 24- to 34-week range, led to a 7-day prevalence of preterm birth of 32.5% versus 19.7% in the largest comparative study of these tests (APOSTEL-1) (see *Chapter 2, Differences between studies in mode of delivery*).

We did not conduct a value-of-information analysis because, as with any model of intervention at birth, there is a degree of structural uncertainty associated with the need to extrapolate from neonatal to lifetime outcomes that would render parameter uncertainty meaningless. Thus, probabilistic sensitivity analysis needs to be considered as exploratory and value of information could not be considered reliable in informing research priorities.

Some of the uncertainties discussed previously may be addressed by the findings of two large (> 1000 participants) ongoing trials of the predictive utility of qfFN (PETRA) and qfFN, PartoSure and Actim Partus (QUIDS 2) in the UK. Until these new data become available, the value of PartoSure is unlikely to be settled. This and other studies may also address a key issue of uncertainty in our analysis, namely whether or not the use of the new tests do affect routine patient management. Our model was limited by the lack of these data and the consequently simplifying assumption that mothers were treated in strict accordance with the test results, which we know does not happen with the current status quo test (fFN at 50 ng/ml).

### 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 ACSs 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 has been done here, modelled on the assumption of full adherence.

## Chapter 10 Conclusions

**A**lthough 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 these 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 or not the use of the biomarker tests for predicting preterm labour led to improved health outcomes. Instead, we used modelling for this purpose.

The limited evidence from DTA studies in patient populations from non-UK countries suggests that Actim Partus may result in a discounted QALY loss of 0.006 per women tested but reduces discounted health care costs by > £30,000 per QALY lost relative to the status quo (the qfFN at 50 ng/ml test). This ratio of savings per QALY lost is larger than that achieved by other options that restrict treatment relative to standard practice (qfFN at 50 ng/ml) (i.e. qfFN at 200 ng/ml and qfFN at 500 ng/ml). In contrast, options that increase access to treatment (i.e. qfFN at 10 ng/ml and the policy of no testing and treat all) increase QALYs and costs and have ICERs in excess of £100,000 relative to qfFN at 50 ng/ml. As for PartoSure, the evidence is inconclusive owing 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 < 30 weeks' gestation without testing may not be cost-effective.

### Suggested research priorities

The primary research priority would be for DTA 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, because population differences between trials would not be an issue. It is probable that these types of studies are currently under way, because some of the index tests are, comparatively speaking, in their infancy of use and to date have been predominantly compared with older (qualitative fFN) or not commonly used (cervical length) tests. We are aware of four ongoing comparative studies, three in the UK and one in France. The publication of these studies will greatly help us compare the DTA between the tests, given that the sample population and hospital treatment protocols will be standardised.

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, which is probably a reflection of the fact that they have included very small samples to reliably measure such outcomes. The importance of this question is highlighted by the emerging evidence documenting the first 48 hours as the target window to optimise ACS treatment.

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

More evidence is required about the side effects of ACS 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.

Improving on the costs of cognitive, respiratory and intestinal neonatal adverse events is also warranted, using electronic records from the Badger data set,<sup>120,121</sup> such as that extracted by the National Neonatal Research Database. This would allow the more precise assessment of 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 DTA and clinical effectiveness. She assisted with the systematic review of economic models and drafting of the background. She provided overall project management. She 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. He contributed to the writing and editing of the report.

**Helen Coelho** contributed to all aspects of the systematic review of DTA and clinical effectiveness. She contributed to the writing and editing of the report.

**Neel Ocean** conducted the review of health-state utility studies and extracted relevant parameters. He 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 DTA and clinical effectiveness. He led the statistical synthesis and meta-analysis. He contributed to the writing and editing of the report.

**David Packman** conducted a review of the cost of adverse events and extracted relevant data. He contributed to model development and to the writing and editing of the report.

**Sophie Dodman** contributed to all aspects of the systematic reviews. She drafted the background section. She assisted with the identification of parameter values for the economic evaluation. She contributed to the writing and editing of the report.

**Chris Cooper** contributed to designing the protocol. He developed and conducted the literature searches for the systematic reviews of DTA, economic evaluations and review of utilities. He 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. He contributed to the writing and editing of the report.

**Tracey Kay** advised on current clinical practice and scientific understanding of preterm labour. She provided individual patient data collected at the Royal Devon and Exeter Hospital. She contributed to the writing and editing of the report.

**Neil Liversedge** advised on current clinical practice and scientific understanding of preterm labour. He contributed to the writing and editing of the report.

**Michelle Parr** advised on current clinical practice and scientific understanding of preterm labour. She contributed to the writing and editing of the report.

**Lisa Knight** advised on current clinical practice and scientific understanding of preterm labour. She provided individual patient data collected at the Royal Devon and Exeter Hospital. She contributed to the writing and editing of the report.

**Chris Hyde** was the project director up to June 2017. He provided DTA advice for both the systematic review and the economic evaluation. He contributed to the writing and editing of the report.

**Andrew Shennan** advised on current clinical practice and scientific understanding of preterm labour. He provided individual patient data from previous research. He contributed to the writing and editing of the report.

**Martin Hoyle** was the project director from June 2017 to completion. He contributed to model checking and advised on model construction. He contributed to the writing and editing of the report.

### Data-sharing statement

Access to the executable economic model developed within this study can be requested and is subject to appropriate agreements being in place. All queries should be submitted to the corresponding author.

### Patient data

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety, and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that it is stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: <https://understandingpatientdata.org.uk/data-citation>.



## References

1. World Health Organization. *Preterm Birth*. URL: [www.who.int/mediacentre/factsheets/fs363/en/](http://www.who.int/mediacentre/factsheets/fs363/en/) (accessed December 2017).
2. 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. *Bull World Health Organ* 2010;**88**:31–8. <https://doi.org/10.2471/BLT.08.062554>
3. Behrman RE, Butler AS. *Mortality and Acute Complications in Preterm Infants*. Washington, DC: National Academies Press; 2007.
4. Moutquin JM. Classification and heterogeneity of preterm birth. *BJOG* 2003;**110**(Suppl. 20):30–3. <https://doi.org/10.1046/j.1471-0528.2003.00021.x>
5. Romero R, Dey SK, Fisher SJ. Preterm labor: one syndrome, many causes. *Science* 2014;**345**:760–5. <https://doi.org/10.1126/science.1251816>
6. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *Lancet* 2008;**371**:75–84. [https://doi.org/10.1016/S0140-6736\(08\)60074-4](https://doi.org/10.1016/S0140-6736(08)60074-4)
7. Halimi AAA, Safari S, Parvareshi Hamrah M. Epidemiology and related risk factors of preterm labor as an obstetrics emergency. *Emerg* 2017;**5**:e3.
8. Goffinet F. Primary predictors of preterm labour. *BJOG* 2005;**112**(Suppl. 1):38–47. <https://doi.org/10.1111/j.1471-0528.2005.00583.x>
9. March of Dimes. *Signs and Symptoms of Preterm Labor*. 2017. URL: [www.marchofdimes.org/complications/signs-and-symptoms-of-preterm-labor.aspx](http://www.marchofdimes.org/complications/signs-and-symptoms-of-preterm-labor.aspx) (accessed October 2017).
10. Ross M, Ridout A, Shennan A. Promising recent advances in preterm birth management. *Front Womens Health* 2016;**1**:27–33.
11. 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**:13–18. [https://doi.org/10.1016/S0002-9378\(97\)70431-9](https://doi.org/10.1016/S0002-9378(97)70431-9)
12. NICE. *DAP40: Biomarker Tests to Help Diagnose Preterm Labour in Women with Intact Membranes – Final Scope*. London: NICE; 2017.
13. Office for National Statistics. *Birth Characteristics in England and Wales: 2016*. London: ONS; 2017.
14. Smith LK, Draper ES, Manktelow BN, Dorling JS, Field DJ. Socioeconomic inequalities in very preterm birth rates. *Arch Dis Child Fetal Neonatal Ed* 2007;**92**:F11–14. <https://doi.org/10.1136/adc.2005.090308>
15. 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. *PGfAR* 2016;**4**(1).
16. 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):2. <https://doi.org/10.1186/1742-4755-10-S1-S2>
17. 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. *Lancet* 2012;**379**:2151–61. [https://doi.org/10.1016/S0140-6736\(12\)60560-1](https://doi.org/10.1016/S0140-6736(12)60560-1)
18. Doyle LW, Ford G, Davis N. Health and hospitalisations after discharge in extremely low birth weight infants. *Semin Neonatol* 2003;**8**:137–45. [https://doi.org/10.1016/S1084-2756\(02\)00221-X](https://doi.org/10.1016/S1084-2756(02)00221-X)

19. 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**:378–84. <https://doi.org/10.1056/NEJM200008103430601>
20. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet* 2008;**371**:261–9. [https://doi.org/10.1016/S0140-6736\(08\)60136-1](https://doi.org/10.1016/S0140-6736(08)60136-1)
21. 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**:155–64. <https://doi.org/10.1097/ANC.0b013e31821baaf4>
22. Melville JM, Moss TJ. The immune consequences of preterm birth. *Front Neurosci* 2013;**7**:79. <https://doi.org/10.3389/fnins.2013.00079>
23. Hodek JM, von der Schulenburg JM, Mittendorf T. Measuring economic consequences of preterm birth - Methodological recommendations for the evaluation of personal burden on children and their caregivers. *Health Econ Rev* 2011;**1**:6. <https://doi.org/10.1186/2191-1991-1-6>
24. NICE. *NG25: Preterm Labour and Birth (Full Guideline)*. London: NICE; 2015.
25. 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. *J Matern Fetal Neonatal Med* 2012;**25**:1690–8. <https://doi.org/10.3109/14767058.2012.657279>
26. Alere Inc. *Alere Actim Partus Test: Instructions for Use*. 2017. URL: [www.medixbiochemica.com/wp-content/uploads/2017/06/K%C3%A4ytt%C3%B6hje-Actim-Partus-OACE31931\\_LR.pdf](http://www.medixbiochemica.com/wp-content/uploads/2017/06/K%C3%A4ytt%C3%B6hje-Actim-Partus-OACE31931_LR.pdf) (accessed January 2019).
27. Parsagen Diagnostics Inc. *PartoSure, Assess the Risk of Preterm Birth: Instructions for Use*. 2015. URL: <https://pdfs.semanticscholar.org/4108/6bda6c63812435f007586a38aa41b53f9d99.pdf> (accessed January 2019).
28. Hologic, Inc. *Rapid fFN 10Q Cassette Kit: Instructions for Use*. 2016. URL: [www.hologic.com/sites/default/files/package-insert/AW-09202-002\\_002\\_02.pdf](http://www.hologic.com/sites/default/files/package-insert/AW-09202-002_002_02.pdf) (accessed January 2019).
29. Hologic, Inc. *QuikCheck fFN Sheet*. URL: [www.hologic.com/sites/default/files/2018-05/AW-05842-002\\_004\\_02.pdf](http://www.hologic.com/sites/default/files/2018-05/AW-05842-002_004_02.pdf) (accessed January 2019).
30. Haas DM, Caldwell DM, Kirkpatrick P, McIntosh JJ, Welton NJ. Tocolytic therapy for preterm delivery: systematic review and network meta-analysis. *BMJ* 2012;**345**:e6226. <https://doi.org/10.1136/bmj.e6226>
31. Elliott JP, Morrison JC, Bofill JA. Risks and benefits of magnesium sulfate tocolysis in preterm labor (PTL). *AIMS Public Health* 2016;**3**:348–56. <https://doi.org/10.3934/publichealth.2016.2.348>
32. Roberts D, Dalziel S. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database Syst Rev* 2006;**3**:CD004454. <https://doi.org/10.1002/14651858.CD004454.pub2>
33. Norman M, Piedvache A, Børch 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**:678–86. <https://doi.org/10.1001/jamapediatrics.2017.0602>
34. Roberts D, Brown J, Medley N, Dalziel SR. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database Syst Rev* 2017;**3**:CD004454. <https://doi.org/10.1002/14651858.CD004454.pub3>
35. 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. *J Paediatr Child Health* 2011;**47**:52–3.

36. Nelson KB, Grether JK. Can magnesium sulfate reduce the risk of cerebral palsy in very low birthweight infants? *Pediatrics* 1995;**95**:263–9.
37. NHS Centre for Reviews and Dissemination. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. 3rd edn. York: Centre for Reviews and Dissemination; 2009.
38. 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. *J Clin Epidemiol* 2011;**64**:602–7. <https://doi.org/10.1016/j.jclinepi.2010.07.006>
39. 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. *Ann Intern Med* 2011;**155**:529–36. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>
40. Deeks J, Bossuyt P, Gatsonis C. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 10*. London: The Cochrane Collaboration; 2010.
41. 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 Fertil* 2015;**43**:640–5. <https://doi.org/10.1016/j.gyobfe.2015.07.002>
42. 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* 2016;**123**:1965–71. <https://doi.org/10.1111/1471-0528.13752>
43. 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. <https://doi.org/10.1016/j.ejogrb.2016.09.018>
44. 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 J Obstet Gynecol* 2017;**7**:11. <https://doi.org/10.4236/ojog.2017.73037>
45. 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**:525–30. <https://doi.org/10.12816/0006491>
46. 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**:279–83. <https://doi.org/10.1007/s00404-008-0703-7>
47. 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**:456–9. <https://doi.org/10.3109/01443615.2010.489162>
48. Brik M, Hernández AI, 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**:268–74. <https://doi.org/10.3109/00016340903443668>
49. 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**:460–5. <https://doi.org/10.1038/jp.2011.133>
50. 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. *J Matern Fetal Neonatal Med* 2011;**24**:1262–6. <https://doi.org/10.3109/14767058.2010.547962>

51. Eroglu D, Yanik F, Oktem M, Zeyneloglu HB, Kuscü E. Prediction of preterm delivery among women with threatened preterm labor. *Gynecol Obstet Invest* 2007;**64**:109–16. <https://doi.org/10.1159/000100120>
52. 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**:165–8. <https://doi.org/10.1016/j.ijgo.2016.01.019>
53. Lembet A, Eroglu D, Ergin T, Kuscü E, Zeyneloglu H, Batioglu S, Haberal A. 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**:706–12. <https://doi.org/10.1034/j.1600-0412.2002.810804.x>
54. 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**:1325–9. <https://doi.org/10.1007/s00404-011-1839-4>
55. 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**:66–72. <https://doi.org/10.1111/j.1447-0756.2008.00833.x>
56. 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**:399–402.
57. 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**:47–50. <https://doi.org/10.1016/j.ijgo.2016.03.030>
58. Vishwekar PS, Chauhan AR, Turakhia N. Prediction of preterm delivery with a novel bedside test. *Int J Reprod Contracept Obstet Gynecol* 2017;**6**:3366–71. <https://doi.org/10.18203/2320-1770.ijrcog20173372>
59. 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;**43**:1263–9. <https://doi.org/10.1111/jog.13366>
60. 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**:473–7. <https://doi.org/10.1515/jpm-2013-0234>
61. 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**:395–402. <https://doi.org/10.1515/jpm-2014-0300>
62. Bruijn MM, Kamphuis EI, Hoesli IM, Martinez de Tejada B, Loccufier AR, Kühnert 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**:793.e1–793.e8. <https://doi.org/10.1016/j.ajog.2016.08.012>
63. 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;**27**:1576–9. <https://doi.org/10.3109/14767058.2013.870550>

64. 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 Health* 2013;**16**:231–50. <https://doi.org/10.1016/j.jval.2013.02.002>
65. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996;**313**:275–83. <https://doi.org/10.1136/bmj.313.7052.275>
66. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Med* 2009;**6**:e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
67. Gibson S, Hezelgrave NL, Shennan AH. The impact of quantitative fetal fibronectin on the clinical management of women symptomatic of preterm labour. *Arch Dis Child Fetal Neonatal Ed* 2014;**99**:A151–5. <https://doi.org/10.1136/archdischild-2014-306576.445>
68. Gibson S. *Implementing Quantitative Fetal Fibronectin into Clinical Practice*. MSc thesis. London: King's College London; 2013.
69. Royal College of Obstetricians and Gynaecologists. *Antenatal Corticosteroids to Reduce Neonatal Morbidity (Green-top 7)*. URL: [www.rcog.org.uk/womens-health/clinical-guidance/antenatal-corticosteroids-prevent-respiratory-distress-syndrome-gree](http://www.rcog.org.uk/womens-health/clinical-guidance/antenatal-corticosteroids-prevent-respiratory-distress-syndrome-gree) (accessed 7 April 2013).
70. Deshpande SN, van Asselt AD, 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 Technol Assess* 2013;**17**(40). <https://doi.org/10.3310/hta17400>
71. Honest H, Forbes CA, Durée KH, Norman G, Duffy SB, Tsourapas A, *et al.* Screening to prevent spontaneous preterm birth: systematic reviews of accuracy and effectiveness literature with economic modelling. *Health Technol Assess* 2009;**13**(43). <https://doi.org/10.3310/hta13430>
72. Chuck A, Nguyen T. *Post Policy Implementation Review (PPIR) of Rapid Fetal Fibronectin Testing for Preterm Labour in Alberta*. Edmonton, AB: Institute of Health Economics; 2015.
73. Boyd KA, Briggs AH, Fenwick E, Norrie J, Stock S. Power and sample size for cost-effectiveness analysis: fFN neonatal screening. *Contemp Clin Trials* 2011;**32**:893–901. <https://doi.org/10.1016/j.cct.2011.07.007>
74. Mozurkewich EL, Naglie G, Krahn MD, Hayashi RH. Predicting preterm birth: a cost-effectiveness analysis. *Am J Obstet Gynecol* 2000;**182**:1589–98. <https://doi.org/10.1067/mob.2000.106855>
75. 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 Obstet Gynecol* 2017;**51**:596–603. <https://doi.org/10.1002/uog.17481>
76. 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**:559–65. <https://doi.org/10.1007/s00404-010-1712-x>
77. Kind P. A comparison of two models for scaling health indicators. *Int J Epidemiol* 1982;**11**:271–5. <https://doi.org/10.1093/ije/11.3.271>
78. 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**:542.e1–7. <https://doi.org/10.1016/j.ajog.2011.09.004>

79. Department of Health and Social Care (DHSC). *NHS Reference Costs 2011–12*. London: DHSC; 2012. URL: [www.gov.uk/government/publications/nhs-reference-costs-financial-year-2011-to-2012](http://www.gov.uk/government/publications/nhs-reference-costs-financial-year-2011-to-2012) (accessed January 2019).
80. Kruse M, Michelsen SI, Flachs EM, Brønnum-Hansen H, Madsen M, Uldall P. Lifetime costs of cerebral palsy. *Dev Med Child Neurol* 2009;**51**:622–8. <https://doi.org/10.1111/j.1469-8749.2008.03190.x>
81. Alvarez MD, Villamil M, Reyes G. Predictive factors in the genesis of intraventricular hemorrhage in premature infants. *P R Health Sci J* 1994;**13**:251–4.
82. Healthcare Commission. *Towards Better Births – A Review of Maternity Services in England*. Commission for Healthcare Audit and Inspection; 2008. URL: [www.healthcarecommission.org.uk](http://www.healthcarecommission.org.uk)
83. 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; Aberdeen: 2007.
84. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010;**13**:509–18. <https://doi.org/10.1111/j.1524-4733.2010.00700.x>
85. British Medical Association (BMA), Royal Pharmaceutical Society of Great Britain (RPS). *British National Formulary No. 62, September 2011*. London: BMA and RPS; 2011. URL: <http://bnf.org/>
86. Villeneuve E, Landa P, Allen M, Spencer A, Prosser S, Gibson A, et al. A framework to address key issues of neonatal service configuration in England: the NeoNet multimethods study. *Health Serv Deliv Res* 2018;**6**(35).
87. 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* 2017;**124**:1567–74. <https://doi.org/10.1111/1471-0528.14545>
88. Sarri G, Davies M, Gholitabar M, Norman JE, Guideline Development Group. Preterm labour: summary of NICE guidance. *BMJ* 2015;**351**:h6283. <https://doi.org/10.1136/bmj.h6283>
89. Ridout A, Watson H, Shennan A. Antenatal corticosteroids in perspective: rationalising current practice. *BJOG* 2016;**123**:1070. <https://doi.org/10.1111/1471-0528.13929>
90. Shennan A, Carter J. *Threatened Preterm Labour: Guideline for Management*. London: Guy's and St Thomas' NHS Foundation Trust; 2017.
91. NICE. *Guide to the Methods of Technology Appraisal 2013*. London: NICE; 2013.
92. 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**:122.e1–6. <https://doi.org/10.1016/j.ajog.2012.10.890>
93. 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* 2017;**356**:j1039. <https://doi.org/10.1136/bmj.j1039>
94. Royal College of Paediatrics and Child Health (RCPCH). *National Neonatal Audit Programme 2016 Annual Report on 2015 Data*. London: RCPCH; 2016.
95. Grant RL. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. *BMJ* 2014;**348**:f7450. <https://doi.org/10.1136/bmj.f7450>
96. 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. *J Neurosurg Pediatr* 2015;**15**:580–8. <https://doi.org/10.3171/2014.11.PEDS14364>

97. Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics Reports. Lifetime costs of cerebral palsy. *Deaths: Final Data for 2004*. 2007;**55**(19).
98. WHO. *WHO Recommendations on Interventions to Improve Preterm Birth Outcomes*. Geneva: WHO; 2015.
99. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;**30**:473–83. <https://doi.org/10.1097/00005650-199206000-00002>
100. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;**37**:53–72. [https://doi.org/10.1016/0168-8510\(96\)00822-6](https://doi.org/10.1016/0168-8510(96)00822-6)
101. Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? *Health Qual Life Outcomes* 2009;**7**:27. <https://doi.org/10.1186/1477-7525-7-27>
102. 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. *J Adolesc Health* 2002;**31**:234–9. [https://doi.org/10.1016/S1054-139X\(01\)00407-4](https://doi.org/10.1016/S1054-139X(01)00407-4)
103. 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 Paediatr* 2013;**102**:607–12. <https://doi.org/10.1111/apa.12217>
104. 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**:S81.
105. Hill PD, Aldag JC. Maternal perceived quality of life following childbirth. *J Obstet Gynecol Neonatal Nurs* 2007;**36**:328–34. <https://doi.org/10.1111/j.1552-6909.2007.00164.x>
106. Couto ER, Couto E, Vian B, Gregório Z, Nomura ML, Zaccaria R, Passini R. Quality of life, depression and anxiety among pregnant women with previous adverse pregnancy outcomes. *Sao Paulo Med J* 2009;**127**:185–9. <https://doi.org/10.1590/S1516-31802009000400002>
107. 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. *J Pediatr* 2016;**179**:68–73.e1. <https://doi.org/10.1016/j.jpeds.2016.08.018>
108. Carroll AE, Downs SM. Improving decision analyses: parent preferences (utility values) for pediatric health outcomes. *J Pediatr* 2009;**155**:21–5.e5. <https://doi.org/10.1016/j.jpeds.2009.01.040>
109. Bastek JA, Langmuir H, Kondapalli LA, Paré E, Adamczak JE, Srinivas SK. Antenatal corticosteroids for late-preterm infants: a decision-analytic and economic analysis. *ISRN Obstet Gynecol* 2012;**2012**:491595. <https://doi.org/10.5402/2012/491595>
110. 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: the EPICure study. *Am J Respir Crit Care Med* 2010;**182**:237–45. <https://doi.org/10.1164/rccm.200912-1806OC>
111. Cooke RW. Health, lifestyle, and quality of life for young adults born very preterm. *Arch Dis Child* 2004;**89**:201–6. <https://doi.org/10.1136/adc.2003.030197>
112. Båtsvik 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 Paediatr* 2015;**104**:948–55. <https://doi.org/10.1111/apa.13069>
113. Lund LK, Vik T, Lydersen S, Løhaugen GC, Skranes J, Brubakk AM, Indredavik MS. Mental health, quality of life and social relations in young adults born with low birth weight. *Health Qual Life Outcomes* 2012;**10**:146. <https://doi.org/10.1186/1477-7525-10-146>

114. 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*. London: NICE; 2010.
115. Vandebussche FP, De Jong-Potjer LC, Stiggelbout AM, Le Cessie S, Keirse MJ. Differences in the valuation of birth outcomes among pregnant women, mothers, and obstetricians. *Birth* 1999;**26**:178–83. <https://doi.org/10.1046/j.1523-536x.1999.00178.x>
116. Parisaei M, Currie J, O’Gorman N, Morris S, David AL. Implementation of foetal fibronectin testing: Admissions, maternal interventions and costs at 1 year. *J Obstet Gynaecol* 2016;**36**:888–92. <https://doi.org/10.3109/01443615.2016.1168374>
117. Curtis L, Burns A. *Unit Costs of Health and Social Care 2016*. Canterbury: Personal Social Services Research Unit, University of Kent; 2016.
118. Gale C, Hay A, Philipp C, Khan R, Santhakumaran S, Ratnavel N. In-utero transfer is too difficult: results from a prospective study. *Early Hum Dev* 2012;**88**:147–50. <https://doi.org/10.1016/j.earlhumdev.2011.07.016>
119. Office for National Statistics. *National Life Tables UK: 2014–2016*. URL: [www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2014to2016](http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2014to2016)
120. Battersby C, Statnikov Y, Santhakumaran S, Gray D, Modi N, Costeloe K, *et al*. The United Kingdom National Neonatal Research Database: a validation study. *PLOS ONE* 2018;**13**:e0201815. <https://doi.org/10.1371/journal.pone.0201815>
121. Villeneuve E, Landa P, Allen M, Spencer A, Prosser S, Gibson A, *et al*. A framework to address key issues of neonatal service configuration in England: the NeoNet multimethods study. *Health Serv Deliv Res* 2018;**6**(35). <https://doi.org/10.3310/hsdr06350>
122. Landry JS, Croitoru D, Jin Y, Schwartzman K, Benedetti A, Menzies D. Health care utilization by preterm infants with respiratory complications in Quebec. *Can Respir J* 2012;**19**:255–60. <https://doi.org/10.1155/2012/606507>
123. 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* 2015;**122**:1495–505. <https://doi.org/10.1111/1471-0528.13515>
124. Gale C, Morris I, Neonatal Data Analysis Unit (NDAU) Steering Board. The UK National Neonatal Research Database: using neonatal data for research, quality improvement and more. *Arch Dis Child Educ Pract Ed* 2016;**101**:216–18. <https://doi.org/10.1136/archdischild-2015-309928>
125. Adams E. *The New NCCMDS, Neonatal HRGs 2016 and Reference Costs: A Guide for Clinicians*. National Casemix Office, NHS England; London: 2016.
126. Mangham LJ, Petrou S, Doyle LW, Draper ES, Marlow N. The cost of preterm birth throughout childhood in England and Wales. *Pediatrics* 2009;**123**:e312–27. <https://doi.org/10.1542/peds.2008-1827>
127. Briggs A, Sculpher M, Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press; 2006.
128. 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. *Obstet Gynecol* 2009;**114**:631–40. <https://doi.org/10.1097/AOG.0b013e3181b47217>
129. 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 meta-analysis. *Am J Obstet Gynecol* 2014;**210**:54.e1–10. <https://doi.org/10.1016/j.ajog.2013.09.004>



130. 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**:1183–91. <https://doi.org/10.1097/AOG.0000000000002367>
131. 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. *J Matern Fetal Neonatal Med* 2017;**26**:1–5. <https://doi.org/10.1080/14767058.2017.1391782>
132. 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;**52**(5):644–9. <https://doi.org/10.1002/uog.18892>
133. Morrison A, Polisen 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. *Int J Technol Assess Health Care* 2012;**28**:138–44. <https://doi.org/10.1017/S0266462312000086>
134. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ* 2013;**185**:E537–44. <https://doi.org/10.1503/cmaj.121286>
135. 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**:2106–16. [https://doi.org/10.1016/S0140-6736\(16\)00350-0](https://doi.org/10.1016/S0140-6736(16)00350-0)
136. Abenhaim HA, Morin L, Benjamin A. Does availability of fetal fibronectin testing in the management of threatened preterm labour affect the utilization of hospital resources? *J Obstet Gynaecol Can* 2005;**27**:689–94. [https://doi.org/10.1016/S1701-2163\(16\)30547-3](https://doi.org/10.1016/S1701-2163(16)30547-3)
137. 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**:581–6. [https://doi.org/10.1016/S0002-9378\(99\)70258-9](https://doi.org/10.1016/S0002-9378(99)70258-9)
138. Berghella V, Saccone G. Fetal fibronectin testing for prevention of preterm birth in singleton pregnancies with threatened preterm labor: a systematic review and meta-analysis of randomized controlled trials. *Am J Obstet Gynecol* 2016;**215**:431–8. <https://doi.org/10.1016/j.ajog.2016.04.038>
139. Lee GT, Burwick R, Zork N, Kjos S. Does the use of fetal fibronectin in an algorithm for preterm labor reduce triage evaluation times? *J Matern Fetal Neonatal Med* 2013;**26**:706–9. <https://doi.org/10.3109/14767058.2012.750291>
140. Duley L, Bennett P. *Tocolysis for Women in Preterm labour. Green-top Guideline No. 1b.* London: Royal College of Obstetricians and Gynaecologists; 2011. URL: [www.rcog.org.uk/files/rcog-corp/GTG1b26072011.pdf](http://www.rcog.org.uk/files/rcog-corp/GTG1b26072011.pdf) (cited 14 February 2012).
141. Roberts D. *Antenatal Corticosteroids to Reduce Neonatal Morbidity and Mortality. Green-top Guideline No. 7.* London: Royal College of Obstetricians and Gynaecologists; 2010. URL: [www.rcog.org.uk/files/rcog-corp/GTG%207.pdf](http://www.rcog.org.uk/files/rcog-corp/GTG%207.pdf) (cited 14 February 2012).
142. Sedgwick P. What is a non-inferiority trial? *BMJ* 2013;**347**:f6853. <https://doi.org/10.1136/bmj.f6853>
143. Healthcare Commission. *Towards Better Births – A Review of Maternity Services in England.* London: Commission for Healthcare Audit and Inspection; 2008.
144. Kenyon SL, Taylor DJ, Tarnow-Mordi W, ORACLE Collaborative Group. Broad-spectrum antibiotics for spontaneous preterm labour: the ORACLE II randomised trial. ORACLE Collaborative Group. *Lancet* 2001;**357**:989–94. [https://doi.org/10.1016/S0140-6736\(00\)04234-3](https://doi.org/10.1016/S0140-6736(00)04234-3)

145. Information Services Division. *Scottish Perinatal Infant Morbidity and Mortality Report 2008*. NHS Scotland; 2009. URL: [www.isdscotland.org/isd/3112.html](http://www.isdscotland.org/isd/3112.html)
146. Macintyre-Beon C, Skeoch C, Jackson L, Boot P, Cameron A. *Perinatal Collaborative Transport Study (CoTS) Final Report*. NHS Quality Improvement Scotland; 2007.
147. Gyetvai K, Hannah ME, Hodnett ED, Ohlsson A. Tocolytics for preterm labor: a systematic review. *Obstet Gynecol* 1999;**94**:869–77. <https://doi.org/10.1097/00006250-199911001-00043>
148. 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**:308–12. <https://doi.org/10.1056/NEJM199207303270503>
149. Crowley PA. Antenatal corticosteroid therapy: a meta-analysis of the randomized trials, 1972 to 1994. *Am J Obstet Gynecol* 1995;**173**:322–35. [https://doi.org/10.1016/0002-9378\(95\)90222-8](https://doi.org/10.1016/0002-9378(95)90222-8)
150. Roos C, Spaanderman ME, Schuit E, Bloemenkamp KW, Bolte AC, Cornette J, et al. Effect of maintenance tocolysis with nifedipine in threatened preterm labor on perinatal outcomes: a randomized controlled trial. *JAMA* 2013;**309**:41–7. <https://doi.org/10.1001/jama.2012.153817>
151. 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). *Am J Perinatol* 2015;**32**:451–60. <https://doi.org/10.1055/s-0034-1390346>
152. 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**:1185–92. <https://doi.org/10.1097/AOG.0000000000000229>
153. Baumgardt M, Bucher HU, Mieth RA, Fauchère JC. Health-related quality of life of former very preterm infants in adulthood. *Acta Paediatr* 2012;**101**:e59–63. <https://doi.org/10.1111/j.1651-2227.2011.02422.x>
154. Berbis J, Einaudi MA, Simeoni MC, Brévaut-Malaty V, Auquier P, d’Ercole C, Gire C. Quality of life of early school-age French children born preterm: a cohort study. *Eur J Obstet Gynecol Reprod Biol* 2012;**162**:38–44. <https://doi.org/10.1016/j.ejogrb.2012.02.006>
155. Bianco F, Cremonesi G, Montiroso 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**:371–2.
156. 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 Hum Dev* 2011;**87**:705–9. <https://doi.org/10.1016/j.earlhumdev.2011.05.011>
157. 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**:e86–93. <https://doi.org/10.1542/peds.2006-2034>
158. 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 Health* 2009;**12**:1124–34. <https://doi.org/10.1111/j.1524-4733.2009.00580.x>
159. 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. *J Paediatr Child Health* 2013;**49**:22–3.

160. van Lunenburg A, van der Pal SM, van Dommelen P, van der Pal-de Bruin KM, Bennebroek Gravenhorst J, Verrips GH. Changes in quality of life into adulthood after very preterm birth and/or very low birth weight in the Netherlands. *Health Qual Life Outcomes* 2013;**11**:51. <https://doi.org/10.1186/1477-7525-11-51>
161. Verrips E, Vogels T, Saigal S, Wolke D, Meyer R, Hoult L, Verloove-Vanhorick SP. Health-related quality of life for extremely low birth weight adolescents in Canada, Germany, and the Netherlands. *Pediatrics* 2008;**122**:556–61. <https://doi.org/10.1542/peds.2007-1043>
162. Wolke D. Born Extremely Low Birth Weight and Health Related Quality of Life into Adulthood. *J Pediatr* 2016;**179**:11–12.e1. <https://doi.org/10.1016/j.jpeds.2016.09.012>
163. 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. *J Pediatr* 2013;**163**:1020–6.e2. <https://doi.org/10.1016/j.jpeds.2013.04.030>
164. 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**:e366–76. <https://doi.org/10.1542/peds.2007-0169>
165. 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. *Arch Pediatr Adolesc Med* 2010;**164**:657–63. <https://doi.org/10.1001/archpediatrics.2010.99>
166. Rautava L, Häkkinen 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. *J Pediatr* 2009;**155**:338–43.e1–3. <https://doi.org/10.1016/j.jpeds.2009.03.061>
167. Lehtonen L, Rautava L, Korvenranta E, Korvenranta H, Peltola M, Hakkinen U. PERFECT preterm infant study. *Annals of Medicine* 2011;**43**:S47–53. <https://doi.org/10.3109/07853890.2011.586359>
168. Dalziel SR, Lim VK, Lambert A, McCarthy D, Parag V, Rodgers A, Harding JE. Psychological functioning and health-related quality of life in adulthood after preterm birth. *Dev Med Child Neurol* 2007;**49**:597–602. <https://doi.org/10.1111/j.1469-8749.2007.00597.x>
169. 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 Qual Life Outcomes* 2016;**14**:56. <https://doi.org/10.1186/s12955-016-0458-y>
170. 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**:100.e1–7. <https://doi.org/10.1016/j.ajog.2016.01.192>
171. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions. *Pharmacoeconomics* 1995;**7**:503–20. <https://doi.org/10.2165/00019053-199507060-00005>
172. Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. *Med Care* 2000;**38**:583–637. <https://doi.org/10.1097/00005650-200006000-00004>
173. Coyle SB. Maternal concern, social support, and health-related quality of life across childhood. *Res Nurs Health* 2011;**34**:297–309. <https://doi.org/10.1002/nur.20438>
174. Pham CT, Crowther CA. Birth outcomes: utility values that postnatal women, midwives and medical staff express. *BJOG* 2003;**110**:121–7. <https://doi.org/10.1046/j.1471-0528.2003.02021.x>
175. 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. *Med Care* 2002;**40**:113–28. <https://doi.org/10.1097/00005650-200202000-00006>

176. 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*. Hamilton, ON: Centre for Health Economics and Policy Analysis, McMaster University; 1998.
177. Saigal S, Feeny 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. *J Pediatr* 1994;**125**:418–25. [https://doi.org/10.1016/S0022-3476\(05\)83289-5](https://doi.org/10.1016/S0022-3476(05)83289-5)
178. 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. *J Pediatr* 1994;**125**:411–17. [https://doi.org/10.1016/S0022-3476\(05\)83288-3](https://doi.org/10.1016/S0022-3476(05)83288-3)
179. 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**:453–9. <https://doi.org/10.1001/jama.1996.03540060029031>
180. 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**:569–74. <https://doi.org/10.1542/peds.105.3.569>
181. Saigal S, Stoskopf B, Pinelli J, Streiner D, Hoult L, Paneth N, Goddeeris J. Self-perceived health-related quality of life of former extremely low birth weight infants at young adulthood. *Pediatrics* 2006;**118**:1140–8. <https://doi.org/10.1542/peds.2006-0119>
182. Kim SH, Kim SO, Lee SI, Jo MW. Deriving a mapping algorithm for converting SF-36 scores to EQ-5D utility score in a Korean population. *Health Qual Life Outcomes* 2014;**12**:145. <https://doi.org/10.1186/s12955-014-0145-9>
183. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. *Health Qual Life Outcomes* 2013;**11**:151. <https://doi.org/10.1186/1477-7525-11-151>
184. Kwon J, Kim SW, Ungar WJ, Tsiplova K, Madan J, Petrou S. A systematic review and meta-analysis of childhood health utilities. *Med Decis Making* 2018;**38**:277–305. <https://doi.org/10.1177/0272989X17732990>
185. Wolke D, Baumann N, Busch B, Bartmann P. Very preterm birth and parents' quality of life 27 years later. *Pediatrics* 2017;**140**:e20171263. <https://doi.org/10.1542/peds.2017-1263>
186. Office for National Statistics. *Pregnancy and Ethnic Factors Influencing Births and Infant Mortality*. London: ONS; 2015.
187. Crump C, Sundquist K, Sundquist J, Winkleby MA. Gestational age at birth and mortality in young adulthood. *JAMA* 2011;**306**:1233–40. <https://doi.org/10.1001/jama.2011.1331>
188. Swamy GK, Ostbye T, Skjaerven R. Association of preterm birth with long-term survival, reproduction, and next-generation preterm birth. *JAMA* 2008;**299**:1429–36. <https://doi.org/10.1001/jama.299.12.1429>
189. Adeza Biomedical. *Adeza Biomedical Fetal Fibronectin Enzyme Immunoassay and Rapid fFN for the TLi<sub>Q</sub> system: Information for Healthcare Providers*. 2003. URL: [http://labmed.ucsf.edu/labmanual/db/resource/FFN\\_Product\\_Insert.pdf](http://labmed.ucsf.edu/labmanual/db/resource/FFN_Product_Insert.pdf) (accessed December 2017).

# Appendix 1 Literature search strategies

## Database searches

### 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 range searched: inception to 24 July 2017.

Date searched: 24 July 2017.

Searcher: Chris Cooper.

Search checked by: Jo Varley-Campbell.

Hits: 916.

### Search strategy

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 plIGFBP-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: not applicable.

File name: PartoSure MEDLINE 916 RIS.txt

### EMBASE

Host: Ovid.

Data parameters: 1974 to 21 July 2017.

Date range searched: inception to 24 July 2017.

Date searched: 24 July 2017.

Searcher: Chris Cooper.

Search checked by: Jo Varley-Campbell.

Hits: 1270.

## Search strategy

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 plIGFBP-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: not applicable.

File name: PartoSure Embase 1270 RIS.

***The Cochrane Library (Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, Database of Abstracts of Reviews of Effects, Health Technology Assessment and NHS Economic Evaluation Database)***

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 EED issue 2 of 4, April 2015).

Date range searched: inception to 24 July 2017.

Date searched: 24 July 2017.

Searcher: Chris Cooper.

Search checked by: Jo Varley-Campbell.

Hits: 159 [note: 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 (totalling 164). NHS EED and DARE were searched as an archive because they have not been updated since 2015]: CDSR 36, DARE 13, CENTRAL 91, HTA 14 and NHS EED 5.

## Search strategy

- #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 plIGFBP-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)

***BioSciences Information Service***

Host: Clarivate Analytics.

Data parameters: 1969–2017.

Date range searched: inception to 24 July 2017.

Date searched: 24 July 2017.

Searcher: Chris Cooper.

Search checked by: Jo Varley-Campbell.

Hits: 806.

### Search strategy

(PartoSure or Parto Sure or PartoSureR or Parto SureR) (2)  
((Placental alpha near/6 test\*) or (PAMG-1)) (25)  
(Actim Partus or Actim PartusR or "insulin-like growth factor-binding protein 1 test" or plIGFBP-1 or  
(IGFBP-1 near/6 test\*)) (80)  
(((Fetal or foetal) near/6 fibronectin\*) or fFN) (716)

Notes: not applicable.

File name: PartoSure BIOSIS 806 RIS.

### Web of Science

Host: Clarivate Analytics.

Data parameters: 1900–2017.

Date range searched: inception to 24 July 2017.

Date searched: 24 July 2017.

Searcher: Chris Cooper.

Search checked by: Jo Varley-Campbell.

Hits: 1358.

### Search strategy

(PartoSure or Parto Sure or PartoSureR or Parto SureR) (3)  
((Placental alpha near/6 test\*) or (PAMG-1)) (45)  
(Actim Partus or Actim PartusR or "insulin-like growth factor-binding protein 1 test" or plIGFBP-1 or  
(IGFBP-1 near/6 test\*)) (124)  
(((Fetal or foetal) near/6 fibronectin\*) or fFN) (1226)

Notes: not applicable.

File name: PartoSure WoS 1358 RIS.

### Cumulative Index to Nursing and Allied Health Literature

Host: EBSCOhost.

Data parameters: 1937–2017.

Date range searched: inception to 24 July 2017.

Date searched: 24 July 2017.

Searcher: Chris Cooper.

Search checked by: Jo Varley-Campbell.

Hits: 258.

### Search strategy

(PartoSure or Parto Sure or PartoSureR or Parto SureR) (4)  
 ((Placental alpha N6 test\*) or (PAMG-1)) (22)  
 (Actim Partus or Actim PartusR or "insulin-like growth factor-binding protein 1 test" or plIGFBP-1 or  
 (IGFBP-1 N6 test\*)) (25)  
 (((Fetal or foetal) N6 fibronectin\*) or fFN) (221)

Notes: not applicable.

File name: PartoSure CINAHL 258 RIS.

### Trial registry searching

Date range searched: inception to 29 August 2017.

Date searched: 29 August 2017.

Searcher: Sophie Dodman.

#### *ClinicalTrials.gov* (<https://clinicaltrials.gov/ct2/home>)

Search terms, enumerated as four different searches:

1. Parto Sure
2. Actim Partus
3. Fetal fibronectin
4. Foetal fibronectin

#### *ISRCTN* ([www.isrctn.com/editAdvancedSearch](http://www.isrctn.com/editAdvancedSearch))

1. Parto Sure
2. Actim Partus
3. Fetal fibronectin
4. Foetal fibronectin

### Web searching

Date searched: 17 September 2017.

Searcher: Chris Cooper.

The first 50 pages were searched in each instance.



## Google

### Search strategy

1. PartoSure
2. PartoSure filetype:pdf
3. "Actim Partus"
4. "Actim Partus" filetype:pdf
5. "Fetal fibronectin"
6. "Fetal fibronectin" filetype:pdf

### Utilities database searches

#### *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 range searched: inception to 11 September 2017.

Date searched: 11 September 2017.

### Search strategy

#	Searches	Results
1	exp Obstetric Labor, Premature/	23,638
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.	188,755
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.	26,742
6	1 or 2 or 3 or 4 or 5	210,020
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 or 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 or sftwenty or shortform twenty or 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 thirtysix or shortform thirty six or short form thirty six or short form thirtysix or short form thirty six).ti,ab,kw.	22,677
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

#	Searches	Results
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/	287,614
22	(health state or health status).ti,ab,kw. or Health status/ or Health status indicators/	127,992
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.	29,499
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	422,994
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	427,211
28	6 and 27	4129
29	limit 28 to english language	3700

## EMBASE

Date range searched: 1974 to 8 September 2017.

Date searched: 11 September 2017.

## Search strategy

#	Searches	Results
1	exp Obstetric Labor, Premature/	40,189
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.	208,972
3	(PROM or PPROM or PROM or PTB).ti,ab,ot.	10,995
4	((Short\$ or reduced or multiple) adj4 gestation\$).ti,ab,ot.	6592
5	(low\$ adj3 birth weight).ti,ab,kw.	33,872
6	1 or 2 or 3 or 4 or 5	238,680
7	(euroqol or euro qol or eq5d or eq 5d or EQ-5D or EQ-5D-Y or EQ-5D-5L).ti,ab,kw.	13,620
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 thirtysix or shortform thirty six or short form thirty six or short form thirtysix or short form thirty six).ti,ab,kw.	33,894
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

#	Searches	Results
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/	464,722
22	(health state or health status).ti,ab,kw. or Health status/ or Health status indicators/	136,100
23	Value of Life/	124,596
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.	42,607
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	726,333
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	732,723
28	6 and 27	8472
29	limit 28 to english language	7634

### NHS Economic Evaluation Database, Issue 2 of 4

Date range searched: inception to April 2015.

Date searched: April 2015.

Date run: 11 September 2017.

### Search strategy

- #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)) (15,146)
- #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 (17,627)

NHS EEDs,  $n = 250$ .

The School of Health and Related Research Health Utilities Database was hand-searched.

## Adverse events database searches

### MEDLINE

Date searched: September 2017.

### Search strategy

---

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 Additional details for quality appraisal for the diagnostic test accuracy review

### 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 preterm and a group of control group women who did not deliver preterm).<sup>41–62</sup> In addition, all included studies avoided inappropriate exclusion of participants in terms of their participant inclusion and exclusion criteria. However, only 5 of the 20 included studies<sup>47,48,50,58,60,61</sup> were rated as having a low risk of bias due to patient selection (four assessing Actim Partus and one assessing PartoSure). These studies reported that eligible women were enrolled into the study consecutively.<sup>47,48,50,58,60,61</sup> For the remaining 15 studies, it was unclear whether or not 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 (see *Table 5*). However, it should be reiterated that in two studies (APOSTEL-1<sup>42,43</sup> and Danti *et al.*<sup>50</sup>), all women included in the final analyses of index test data had a transvaginal cervical length measurement of  $\leq 30$  mm, which would probably increase the prevalence of preterm birth in these studies.

### 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<sup>42,43</sup> evaluated both Actim Partus and qfFN and the study by Hadzi-Lega *et al.*<sup>44</sup> evaluated both Actim Partus and PartoSure. The other 18 studies all evaluated only one of the index tests of interest, with three of the studies evaluating only PartoSure,<sup>41,59–61</sup> the Buijn study only evaluating qfFN<sup>62</sup> and the remaining 14 studies evaluating only Actim Partus.

In all studies included in the review, except for the APOSTEL-1 study<sup>42,43</sup> and the study by Cooper *et al.*,<sup>49</sup> the risk that the conduct and interpretation of the test could have introduced bias was rated as low (see *Table 5*). This is because all studies either clearly reported prespecified thresholds for the test (qfFN) or used a test with a standardised threshold (Actim Partus or PartoSure), and owing to the timing of the tests, all studies other than APOSTEL-1<sup>42,43</sup> and the study by Cooper *et al.*<sup>49</sup> interpreted the index tests without knowledge of the reference standard (the tests were conducted before the occurrence of preterm birth). In both the APOSTEL-1 study and the study by Cooper *et al.*,<sup>49</sup> frozen samples were used (see *Chapter 2, Frozen samples*). In both studies, although samples were collected prior to the assessment of the reference standard, the index tests (Actim Partus and qfFN in APOSTEL-1<sup>42,43</sup> and Actim Partus in Cooper *et al.*<sup>49</sup>) were interpreted after the assessment of the reference standard. For these index tests, it is unclear whether or not these interpretations were made blind to whether or not preterm birth had occurred.<sup>42,43,49</sup> 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, in the studies assessing more than one test (APOSTEL-1<sup>42,43</sup> and that by Hadzi-Lega *et al.*<sup>44</sup>) 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 owing 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 qfFN test does not require any subjective judgement of the test result, whereas both the Actim Partus and PartoSure tests 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 qfFN test when no subjective interpretation is required. For this reason, in the APOSTEL-1 study, the Actim Partus test was conducted before the qfFN test.<sup>42,43</sup> However, additional tests were also carried out in the APOSTEL-1 study (qualitative fFN and cervical length measurement) and these were conducted before the index tests; it is unclear whether or not the index tests were interpreted blind to the results from the qualitative fFN test or the cervical length measurement. It is therefore unclear whether or not any 'cross-contamination' of results may have taken place between the two tests, although, again, owing to the nature of the tests the scope for such bias is very limited. In the study by Hadzi-Lega *et al.*,<sup>44</sup> in which both the PartoSure and Actim Partus tests were conducted, it was unclear which test was carried out first, or indeed whether or not the tests were carried out in a predetermined order.<sup>44</sup> 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<sup>47,49,52</sup> and one assessing PartoSure,<sup>59</sup> it was reported that the index test was interpreted blind to the results of transvaginal cervical length measurements<sup>47,49,52</sup> or qualitative fFN.<sup>59</sup> In the studies by Azlin *et al.*,<sup>47</sup> Bolotskikh *et al.*<sup>59</sup> and Goyal *et al.*,<sup>52</sup> this was because the cervical length measurement was conducted after the index test and in the study by Cooper *et al.*<sup>59</sup> it was stated that the Actim Partus test was conducted blind to the qualitative fFN measurement. In an additional study,<sup>48</sup> it was stated that the sample was collected before the cervical length measurement and that the Actim Partus test was conducted immediately, so it is likely that the index test was interpreted before the cervical length measurement. In a further seven studies,<sup>50,51,54,56,57,60-62</sup> additional tests or assessments (qualitative fFN and/or cervical length measurement) were reported, but it was unclear whether or not the index tests were interpreted blind to the results of these additional tests. In the remaining six studies,<sup>41,45,46,53,55,58</sup> there was no reported use of other tests or clinical assessments.

In all of the studies included in the review, except for the APOSTEL-1 study<sup>42,43</sup> and the study by Ting *et al.*,<sup>56</sup> there were no concerns that the conduct or interpretation of either of the index tests was different from the review question (see *Table 5*). However, in APOSTEL-1<sup>42,43</sup> and the study by Ting *et al.*,<sup>56</sup> the test was conducted in a way that may have had an impact on the results because samples were taken from the posterior fornix of the vagina rather than the external cervical os.

## Reference standard

In all of the included studies, the reference standard was whether or not preterm birth took place within 48 hours and/or within 7 days.<sup>41-62</sup> For this reason, 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 (preterm birth), did not match the review question (see *Table 5*).

In addition, because the reference standard was the occurrence of preterm birth rather than a diagnostic test, no assessment was made regarding whether or not 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). For this reason, there were no concerns in any of the included studies regarding whether the reference standard, its conduct or its interpretation could have introduced bias (see *Table 5*).

## Flow and timing

Of course, in all of the included studies, all participants 'received' a reference standard.<sup>41–62</sup> Because the reference standard was the occurrence or otherwise of preterm delivery by 48 hours and/or by 7 days, this was the same for all participants in all of the included studies (see *Table 5*).

It is important to note that because the index tests in this review (PartoSure, Actim Partus and qfFN) 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, although indicated, may inflate false-positive results and deflate true-positive results.

For all but two of the included studies (APOSTEL-1<sup>42,43</sup> and Cooper *et al.*<sup>49</sup>), the timing of the index test was as per the manufacturers' instructions. For both of these studies, 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.<sup>42,43,49</sup> However, this is unlikely to have had a large impact on the test results (see *Chapter 2, Frozen samples*, for further details).

Related to this, and outside 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 or not 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 *et al.*<sup>47</sup> and Riboni *et al.*<sup>54</sup>), although in the study by Riboni *et al.*<sup>54</sup> it was stated that clinical management was not altered by this knowledge. In six of the studies,<sup>42–44,49,50,60–62</sup> it was reported that clinical management personnel were unaware of the index tests (management was based on qualitative fFN results and/or cervical length). In four of the studies,<sup>51,53,55,56</sup> clinical management personnel were blinded to the results of all tests reported in the study and, therefore, it was unclear on what basis clinical decisions were made. In the remaining eight studies,<sup>41,45,46,48,52,57–59</sup> it was not clearly reported whether or not clinical personnel were blinded to the index test results.

Regarding missing data, it was clear that in 10 of the included studies<sup>41–45,48,52,56,57,60–62</sup> some of the participants were excluded from the analysis (see *Table 5*). This may lead to bias because it was not clear from the study reports whether or not the women who were not included systematically differed from those whose data were analysed. In fact, only eight studies<sup>47,49–51,53,55,58,59</sup> clearly specify that data were analysed from all participants who received tests, although it should be noted that, in the study by Tanir *et al.*,<sup>55</sup> two participants with failed tests were coded as test positives.





## Appendix 3 Supplementary citations and data

### Included systematic reviews

Corabian P. The Actim™ Partus versus the TLI<sub>Q</sub>® System as rapid response tests to aid in diagnosing preterm labour in symptomatic women (Structured abstract). *Health Technol Assess* 2008;(4).

URL: <http://onlinelibrary.wiley.com/doi/10.1002/hta.32008100231/frame.html> (accessed September 2017).

Corabian P, Guo B, Chojecki D, Wodinski L, Wanke M, Nguyen T, Chuck A. Post policy implementation review (PPIR) of rapid fetal fibronectin testing for preterm labour in Alberta (Structured abstract). *Health Technol Assess* 2015;(4). URL: <http://onlinelibrary.wiley.com/doi/10.1002/hta.32015001020/frame.html> (accessed September 2017).

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. *Am J Obstet Gynecol* 2016;**214**:57–73. <https://doi.org/10.1016/j.ajog.2015.06.060>

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* 2002;**325**:301.

Leitch H, Kaider A. Fetal fibronectin – how useful is it in the prediction of preterm birth? *BJOG* 2003;**110**(Suppl. 20):66–70.

Lucaroni F, Morciano L, Rizzo G, D'Antonio F, Buonomo E, Palombi L, et al. Biomarkers for predicting spontaneous preterm birth: an umbrella systematic review. *J Matern Fetal Neonatal Med* 2017;**31**:1–9.

Menon R, Torloni MR, Voltolini C, Torricelli M, Merialdi M, Betrán AP, et al. Biomarkers of spontaneous preterm birth: an overview of the literature in the last four decades. *Reprod Sci* 2011;**18**:1046–70. <https://doi.org/10.1177/1933719111415548>

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. *Am J Obstet Gynecol* 2007;**197**:S198-S.

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. *Obstet Gynecol* 2009;**114**:631–40. <https://doi.org/10.1097/AOG.0b013e3181b47217>

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. *Eur J Obstet Gynecol Reprod Biol* 2007;**133**:134–42.

Vis JY, Wilms FF, Oudijk MA, Bossuyt PM, van der Post JA, Grobman WA, Mol BW. Why were the results of randomized trials on the clinical utility of fetal fibronectin negative? A systematic review of their study designs. *Am J Perinatol* 2011;**28**:145–50. <https://doi.org/10.1055/s-0030-1263297>

## Included citations

TABLE 27 Included citations

Reference	Full text or abstract
Abo El-Ezz AE, Askar AE. Predictive value of phosphorylated insulin-like growth factor binding protein-1 (PIGFBP-1) (bedside test) in preterm labor. <i>J Egypt Soc Parasitol</i> 2014; <b>44</b> :525–30	FT
Altinkaya O, Gungor T, Ozat M, Danisman N, Mollamahmutoglu L. Cervical phosphorylated insulin-like growth factor binding protein-1 in prediction of preterm delivery. <i>Arch Gynecol Obstet</i> 2009; <b>279</b> :279–83. <a href="https://doi.org/10.1007/s00404-008-0703-7">https://doi.org/10.1007/s00404-008-0703-7</a>	FT
Azlin MI, Bang HK, An LJ, Mohamad SN, Mansor NA, Yee BS, <i>et al.</i> Role of pIIGFBP-1 and ultrasound cervical length in predicting pre-term labour. <i>J Obstet Gynaecol</i> 2010; <b>30</b> :456–9. <a href="https://doi.org/10.3109/01443615.2010.489162">https://doi.org/10.3109/01443615.2010.489162</a>	FT
Azlin MN, Kee BH, Low JA, Mohamad SN, Mansor NA, Bee SY, <i>et al.</i> Role of pIIGFBP-1 and cervical length in predicting preterm labor. <i>J Maternal Fetal Neonatal Med</i> 2010; <b>23</b> :100	Ab
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. <i>J Obstet Gynaecol Res</i> 2017; <b>43</b> :1263–9. <a href="https://doi.org/10.1111/jog.13366">https://doi.org/10.1111/jog.13366</a>	FT
Brik M, Hernández AI, Pedraz CC, Perales A. Phosphorylated insulin-like growth factor binding protein-1 and cervical measurement in women with threatening preterm birth. <i>Acta Obstet Gynecol Scand</i> 2010; <b>89</b> :268–74. <a href="https://doi.org/10.3109/00016340903443668">https://doi.org/10.3109/00016340903443668</a>	FT
Bruijn M, Kamphuis E, Hoesli I, de Tejada BM, Loccufier A, Jacquemyn Y, <i>et al.</i> 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). <i>Am J Obstet Gynecol</i> 2015; <b>212</b> (Suppl. S):S190–1	Ab
Bruijn M, Kamphuis E, Hoesli I, de Tejada BM, Loccufier A, Jacquemyn Y, <i>et al.</i> 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). <i>Am J Obstet Gynecol</i> 2015; <b>212</b> :S191–2	Ab
Bruijn M, van Baaren G-J, Vis J, van Straalen J, Wilms F, Oudijk M, <i>et al.</i> 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. <i>Am J Obstet Gynecol</i> 2014; <b>210</b> (Suppl. S):S364	Ab
Bruijn M, van Baaren G-J, Vis J, van Straalen J, Wilms F, Oudijk M, <i>et al.</i> 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. <i>Am J Obstet Gynecol</i> 2014; <b>210</b> (Suppl. S):S363–4	Ab
Bruijn M, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM, <i>et al.</i> Quantitative fetal fibronectin testing in combination with cervical length measurement in the prediction of spontaneous preterm delivery in symptomatic women. <i>BJOG</i> 2016; <b>123</b> :1965–71. <a href="https://doi.org/10.1111/1471-0528.13752">https://doi.org/10.1111/1471-0528.13752</a>	FT
Bruijn MM, Kamphuis EI, Hoesli IM, Martinez de Tejada B, Loccufier AR, Kühnert M, <i>et al.</i> The predictive value of quantitative fibronectin testing in combination with cervical length measurement in symptomatic women. <i>Am J Obstet Gynecol</i> 2016; <b>215</b> :793.e1–793.e8	FT
Bruijn MM, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM, <i>et al.</i> 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. <i>Eur J Obstet Gynecol Reprod Biol</i> 2016; <b>206</b> :220–4	FT
Cooper S, Lange I, Wood S, Tang S, Miller L, Ross S. Diagnostic accuracy of rapid pIIGFBP-I assay for predicting preterm labor in symptomatic patients. <i>J Perinatal</i> 2012; <b>32</b> :460–5. <a href="https://doi.org/10.1038/jp.2011.133">https://doi.org/10.1038/jp.2011.133</a>	FT
Danti L, Prefumo F, Lojacono A, Corini S, Testori A, Frusca T. The combination of short cervical length and pIIGFBP-1 in the prediction of preterm delivery in symptomatic women. <i>J Matern Fetal Neonatal Med</i> 2011; <b>24</b> :1262–6. <a href="https://doi.org/10.3109/14767058.2010.547962">https://doi.org/10.3109/14767058.2010.547962</a>	FT
Eroglu D, Yanik F, Oktem M, Zeyneloglu HB, Kuscu E. Prediction of preterm delivery among women with threatened preterm labor. <i>Gynecol Obstet Invest</i> 2007; <b>64</b> :109–16	FT

TABLE 27 Included citations (continued)

Reference	Full text or abstract
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. <i>Int J Gynaecol Obstet</i> 2016; <b>134</b> :165–8. <a href="https://doi.org/10.1016/j.ijgo.2016.01.019">https://doi.org/10.1016/j.ijgo.2016.01.019</a>	FT
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. <i>Ultrasound Obstet Gynecol</i> 2016; <b>48</b> :345–6	Ab
Hadzi-Lega M, Hellmeyer L, Hanns H, Josefine M, Poposka A, Daneva Markova A. Comparison of PartoSure (PAMG-1) and Actim Partus (phIGFBP-1) for the prediction of preterm delivery in patients with preterm labor and a short cervix. <i>J Maternal Fetal Neonatal Med</i> 2016; <b>29</b> :55	Ab
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. <i>Open J Obstet Gynecol</i> 2017; <b>7</b> :11	FT
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. <i>Am J Obstet Gynecol</i> 2001; <b>185</b> (Suppl. 6):S147	Ab
Lembet A, Eroglu D, Ergin T, Kuscu E, Zeyneloglu H, Batioglu S, Haberal A. New rapid bed-side test to predict preterm delivery: phosphorylated insulin-like growth factor binding protein-1 in cervical secretions. <i>Acta Obstet Gynecol Scand</i> 2002; <b>81</b> :706–12	FT
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. Abstract number 1731. Moscow: 11th World Congress of Perinatal Medicine; 2013	Ab
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. <i>J Perinat Med</i> 2014; <b>42</b> :473–7. <a href="https://doi.org/10.1515/jpm-2013-0234">https://doi.org/10.1515/jpm-2013-0234</a>	FT
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. <i>J Perinat Med</i> 2015; <b>43</b> :395–402. <a href="https://doi.org/10.1515/jpm-2014-0300">https://doi.org/10.1515/jpm-2014-0300</a>	FT
Riboni F, Vitulo A, Dell'avano 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. <i>Arch Gynecol Obstet</i> 2011; <b>284</b> :1325–9. <a href="https://doi.org/10.1007/s00404-011-1839-4">https://doi.org/10.1007/s00404-011-1839-4</a>	FT
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. <i>J Obstet Gynaecol Res</i> 2009; <b>35</b> :66–72. <a href="https://doi.org/10.1111/j.1447-0756.2008.00833.x">https://doi.org/10.1111/j.1447-0756.2008.00833.x</a>	FT
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. <i>Ann Acad Med Singap</i> 2007; <b>36</b> :399–402	FT
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. <i>Int J Gynaecol Obstet</i> 2016; <b>135</b> :47–50. <a href="https://doi.org/10.1016/j.ijgo.2016.03.030">https://doi.org/10.1016/j.ijgo.2016.03.030</a>	FT
Vishwekar PS, Chauhan AR, Turakhia N. Prediction of preterm delivery with a novel bedside test. <i>Int J Reprod Contracept Obstet Gynecol</i> 2017; <b>6</b> :3366–71	FT
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.] <i>Gynecol Obstet Fertil</i> 2015; <b>43</b> :640–5. <a href="https://doi.org/10.1016/j.gyobfe.2015.07.002">https://doi.org/10.1016/j.gyobfe.2015.07.002</a>	FT
Ab, abstract; FT, full text.	

## Alere Inc. (Actim Partus) submitted citations

TABLE 28 Alere Inc. (Actim Partus) submitted citations

Citation	Reason for exclusion	Further detail
Adeyemi O, Osoba L. The role of phosphorylated insulin-like growth factor binding protein-1 in predicting pre-term labour in twin pregnancies. <i>J Obstet Gynaecol</i> 2010; <b>30</b> :571–3. <a href="https://doi.org/10.3109/01443615.2010.494203">https://doi.org/10.3109/01443615.2010.494203</a>	Population	Twin pregnancies
Akercan F, Kazandi M, Sendag F, Cirpan T, Mgoyi L, Terek MC, Sagol S. Value of cervical phosphorylated insulinlike growth factor binding protein-1 in the prediction of preterm labor. <i>J Reprod Med</i> 2004; <b>49</b> :368–72	Comparator	Outcome at 37 weeks
Altinkaya O, Gungor T, Ozat M, Danisman N, Mollamahmutoglu L. Cervical phosphorylated insulin-like growth factor binding protein-1 in prediction of preterm delivery. <i>Arch Gynecol Obstet</i> 2009; <b>279</b> :279–83. <a href="https://doi.org/10.1007/s00404-008-0703-7">https://doi.org/10.1007/s00404-008-0703-7</a>	Included	
Azlin MI, Bang HK, An LJ, Mohamad SN, Mansor NA, Yee BS, <i>et al.</i> Role of pIIGFBP-1 and ultrasound cervical length in predicting pre-term labour. <i>J Obstet Gynaecol</i> 2010; <b>30</b> :456–9. <a href="https://doi.org/10.3109/01443615.2010.489162">https://doi.org/10.3109/01443615.2010.489162</a>	Included	
Balić D, Latifagić A, Hudić I. Insulin-like growth factor-binding protein-1 (IGFBP-1) in cervical secretions as a predictor of preterm delivery. <i>J Matern Fetal Neonatal Med</i> 2008; <b>21</b> :297–300. <a href="https://doi.org/10.1080/14767050802037613">https://doi.org/10.1080/14767050802037613</a>	Population	Asymptomatic population
Bittar RE, da Fonseca EB, de Carvalho MH, 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. <i>Ultrasound Obstet Gynecol</i> 2007; <b>29</b> :562–7. <a href="https://doi.org/10.1002/uog.3989">https://doi.org/10.1002/uog.3989</a>	Population	Asymptomatic population
Brik M, Hernández AI, Pedraz CC, Perales A. Phosphorylated insulin-like growth factor binding protein-1 and cervical measurement in women with threatening preterm birth. <i>Acta Obstet Gynecol Scand</i> 2010; <b>89</b> :268–74. <a href="https://doi.org/10.3109/00016340903443668">https://doi.org/10.3109/00016340903443668</a>	Included	
Bruijn MM, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM, <i>et al.</i> 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. <i>Eur J Obstet Gynecol Reprod Biol</i> 2016; <b>206</b> :220–4	Included	
Cooper S, Lange I, Wood S, Tang S, Miller L, Ross S. Diagnostic accuracy of rapid pIIGFBP-1 assay for predicting preterm labor in symptomatic patients. <i>J Perinatol</i> 2012; <b>32</b> :460–5. <a href="https://doi.org/10.1038/jp.2011.133">https://doi.org/10.1038/jp.2011.133</a>	Included	
Danti L, Prefumo F, Lojaco A, Corini S, Testori A, Frusca T. The combination of short cervical length and pIIGFBP-1 in the prediction of preterm delivery in symptomatic women. <i>J Matern Fetal Neonatal Med</i> 2011; <b>24</b> :1262–6. <a href="https://doi.org/10.3109/14767058.2010.547962">https://doi.org/10.3109/14767058.2010.547962</a>	Included	
Dögl 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. <i>Acta Obstet Gynecol Scand</i> 2011; <b>90</b> :57–62. <a href="https://doi.org/10.1111/j.1600-0412.2010.01018.x">https://doi.org/10.1111/j.1600-0412.2010.01018.x</a>	Population	Post-term pregnancies
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. <i>Am J Perinatol</i> 2005; <b>22</b> :305–9. <a href="https://doi.org/10.1055/s-2005-870895">https://doi.org/10.1055/s-2005-870895</a>	Population	Not clear if multifetal, also outcome at 35 weeks

TABLE 28 Alere Inc. (Actim Partus) submitted citations (continued)

Citation	Reason for exclusion	Further detail
Eroglu D, Yanik F, Oktem M, Zeyneloglu HB, Kuscü E. Prediction of preterm delivery among women with threatened preterm labor. <i>Gynecol Obstet Invest</i> 2007; <b>64</b> :109–16	Included	
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. <i>J Perinat Med</i> 2015; <b>43</b> :545–51. <a href="https://doi.org/10.1515/jpm-2014-0275">https://doi.org/10.1515/jpm-2014-0275</a>	Comparator	Outcome at 14 days
Kekki M, Kurki T, Kärkkäinen T, Hiilesmaa V, Paavonen J, Rutanen EM. Insulin-like growth factor-binding protein-1 in cervical secretion as a predictor of preterm delivery. <i>Acta Obstet Gynecol Scand</i> 2001; <b>80</b> :546–51	Outcome	Unclear what time reference is, also includes multifetal
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. <i>Lancet</i> 1999; <b>353</b> :1494	Study design	Letter
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. <i>J Obstet Gynaecol</i> 2012; <b>32</b> :132–4. <a href="https://doi.org/10.3109/01443615.2011.637649">https://doi.org/10.3109/01443615.2011.637649</a>	Population	Asymptomatic
Kosinska-Kaczynska K, Bomba-Opon D, Bobrowska K, Kozłowski S, Brawura-Biskupski-Samaha R, Szymusik I, et al. Phosphorylated IGFBP-1 in predicting successful vaginal delivery in post-term pregnancy. <i>Arch Gynecol Obstet</i> 2015; <b>292</b> :45–52. <a href="https://doi.org/10.1007/s00404-014-3577-x">https://doi.org/10.1007/s00404-014-3577-x</a>	Population	Post-term population
Kwek K, Khi C, Ting HS, Yeo GS. Evaluation of a bedside test for phosphorylated insulin-like growth factor binding protein-1 in preterm labour. <i>Ann Acad Med Singap</i> 2004; <b>33</b> :780–3	Population	Unclear if includes multifetal
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. <i>Med Glas</i> 2008; <b>5</b> :121–4	Outcome	Unclear what time reference is
Lembet A, Eroglu D, Ergin T, Kuscü E, Zeyneloglu H, Batioglu S, Haberal A. New rapid bed-side test to predict preterm delivery: phosphorylated insulin-like growth factor binding protein-1 in cervical secretions. <i>Acta Obstet Gynecol Scand</i> 2002; <b>81</b> :706–12	Included	
Mešić Đogić L, Mičić D, Omeragić F, Kovač R, Fazlagić S. IGFBP-1 marker of cervical ripening and predictor of preterm birth. <i>Med Glas</i> 2016; <b>13</b> :118–24. <a href="https://doi.org/10.17392/856-16">https://doi.org/10.17392/856-16</a>	Population	Asymptomatic population
Nuutila M, Hiilesmaa V, Kärkkäinen T, Ylikorkala O, Rutanen EM. Phosphorylated isoforms of insulin-like growth factor binding protein-1 in the cervix as a predictor of cervical ripeness. <i>Obstet Gynecol</i> 1999; <b>94</b> :243–9	Population	Term population
Park OR, Kim JK, Chang BS, Kim HJ, Kim TS, Park IS, et al. Usefulness of phosphorylated insulin-like growth factor binding protein-1 for prediction of preterm delivery. <i>Korean J Obstet Gynecol</i> 2003; <b>46</b> :1378–83	Language	
Paternoster DM, Muresan D, Vitulo A, Serena A, Battagliarin G, Dell'avano M, Nicolini U. Cervical pHIGFBP-1 in the evaluation of the risk of preterm delivery. <i>Acta Obstet Gynecol Scand</i> 2007; <b>86</b> :151–5	Outcome	Outcome at 34 weeks
Rahkonen L. Prediction of pre-term delivery with phosphorylated insulin-like growth factor-binding protein-1. <i>Eur J Obstet Gynecol</i> 2011; <b>6</b> :3–7	Study design	Review
Riboni F, Vitulo A, Dell'avano 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. <i>Arch Gynecol Obstet</i> 2011; <b>284</b> :1325–9. <a href="https://doi.org/10.1007/s00404-011-1839-4">https://doi.org/10.1007/s00404-011-1839-4</a>	Included	

continued

TABLE 28 Alere Inc. (Actim Partus) submitted citations (continued)

Citation	Reason for exclusion	Further detail
Riboni F, Vitulo A, Plebani M, Dell'avano M, Battagliarin G, Paternoster D. Combination of biochemical markers in predicting pre-term delivery. <i>Arch Gynecol Obstet</i> 2012; <b>285</b> :61–6. <a href="https://doi.org/10.1007/s00404-011-1915-9">https://doi.org/10.1007/s00404-011-1915-9</a>	Population	Asymptomatic
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.] <i>Rev Bras Ginecol Obstet</i> 2013; <b>35</b> :394–400	Population	Asymptomatic
Shine BK, Kim SJ, Park WI, Kim JO, Kim DW, Hong SY, Yoon HS. Insulin-like growth factor-binding protein-1 in cervical secretion as a predictor of preterm delivery. <i>Korean J Obstet Gynecol</i> 2001; <b>44</b> :2250–6	Language	
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. <i>J Obstet Gynaecol Res</i> 2009; <b>35</b> :66–72. <a href="https://doi.org/10.1111/j.1447-0756.2008.00833.x">https://doi.org/10.1111/j.1447-0756.2008.00833.x</a>	Included	
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. <i>Ann Acad Med Singap</i> 2007; <b>36</b> :399–402	Included	
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. <i>Int J Gynaecol Obstet</i> 2016; <b>135</b> :47–50. <a href="https://doi.org/10.1016/j.ijgo.2016.03.030">https://doi.org/10.1016/j.ijgo.2016.03.030</a>	Included	
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. <i>BJOG</i> 2017; <b>124</b> :1274–83. <a href="https://doi.org/10.1111/1471-0528.14175">https://doi.org/10.1111/1471-0528.14175</a>	Population	Term women, labour induction

## Hologic, Inc. (fFN) submitted citations

TABLE 29 Hologic, Inc. (fFN) submitted citations

Citation	Reason for exclusion	Further detail
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. <i>J Obstet Gynaecol</i> 2013; <b>33</b> :414–15. <a href="https://doi.org/10.3109/01443615.2013.772129">https://doi.org/10.3109/01443615.2013.772129</a>	Study design	Case study
Abbott DS, Hezelgrave NL, Seed PT, Norman JE, David AL, Bennett PR, <i>et al.</i> Quantitative fetal fibronectin to predict preterm birth in asymptomatic women at high risk. <i>Obstet Gynecol</i> 2015; <b>125</b> :1168–76. <a href="https://doi.org/10.1097/AOG.0000000000000754">https://doi.org/10.1097/AOG.0000000000000754</a>	Population	Asymptomatic population
Abbott DS, Radford SK, Seed PT, Tribe RM, Shennan AH. Evaluation of a quantitative fetal fibronectin test for spontaneous preterm birth in symptomatic women. <i>Am J Obstet Gynecol</i> 2013; <b>208</b> :122.e1–6. <a href="https://doi.org/10.1016/j.ajog.2012.10.890">https://doi.org/10.1016/j.ajog.2012.10.890</a>	Comparator	Outcome at 14 days
Anderson-Knight HE, Hezelgrave NL, Shennan AH. Spontaneous resolution of a midtrimester dilated cervix with expectant management guided by quantitative foetal fibronectin results. <i>J Obstet Gynaecol</i> 2015; <b>35</b> :766–7. <a href="https://doi.org/10.3109/01443615.2015.1006597">https://doi.org/10.3109/01443615.2015.1006597</a>	Study design	Case study

TABLE 29 Hologic, Inc. (fFN) submitted citations (continued)

Citation	Reason for exclusion	Further detail
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? <i>Eur J Obstet Gynecol Reprod Biol</i> 2009; <b>146</b> :180–3. <a href="https://doi.org/10.1016/j.ejogrb.2009.06.021">https://doi.org/10.1016/j.ejogrb.2009.06.021</a>	Comparator	Outcome at 14 days
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. <i>J Matern Fetal Neonatal Med</i> 2011; <b>24</b> :928–32. <a href="https://doi.org/10.3109/14767058.2010.535872">https://doi.org/10.3109/14767058.2010.535872</a>	Population	Asymptomatic population
Bolt LA, Morrison K, Shennan AH. The use of fetal fibronectin testing and cervical length measurement in the prediction of delivery of triplet pregnancies. <i>Eur J Obstet Gynecol Reprod Biol</i> 2012; <b>164</b> :236–7. <a href="https://doi.org/10.1016/j.ejogrb.2012.06.012">https://doi.org/10.1016/j.ejogrb.2012.06.012</a>	Population	Triplet pregnancies
Bruijn M, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM, <i>et al.</i> Quantitative fetal fibronectin testing in combination with cervical length measurement in the prediction of spontaneous preterm delivery in symptomatic women. <i>BJOG</i> 2016; <b>123</b> :1965–71. <a href="https://doi.org/10.1111/1471-0528.13752">https://doi.org/10.1111/1471-0528.13752</a>	Included	
Bruijn MM, Kamphuis EI, Hoesli IM, Martinez de Tejada B, Loccufer AR, Kühnert M, <i>et al.</i> The predictive value of quantitative fibronectin testing in combination with cervical length measurement in symptomatic women. <i>Am J Obstet Gynecol</i> 2016; <b>215</b> :793.e1–793.e8	Included	
Centra M, Coata G, Picchiassi E, Alfonsi L, Meniconi S, Bini V, <i>et al.</i> Evaluation of quantitative fFn test in predicting the risk of preterm birth. <i>J Perinat Med</i> 2017; <b>45</b> :91–8. <a href="https://doi.org/10.1515/jpm-2015-0414">https://doi.org/10.1515/jpm-2015-0414</a>	Comparator	Outcome at 14 days
Fiorini F, Isted A, Hezelgrave NL, Shennan AH. Quantitative fetal fibronectin predicts preterm birth in women with bulging fetal membranes. <i>Eur J Obstet Gynecol Reprod Biol</i> 2016; <b>203</b> :127–31. <a href="https://doi.org/10.1016/j.ejogrb.2016.05.046">https://doi.org/10.1016/j.ejogrb.2016.05.046</a>	Comparator	Outcome at 14 days
Foster C, Shennan AH. Fetal fibronectin as a biomarker of preterm labor: a review of the literature and advances in its clinical use. <i>Biomark Med</i> 2014; <b>8</b> :471–84. <a href="https://doi.org/10.2217/bmm.14.28">https://doi.org/10.2217/bmm.14.28</a>	Study design	Literature review
Gibson S, Hezelgrave NL, Shennan AH. Management of vasa praevia: a potential role for cervical length and quantitative fetal fibronectin measurement. <i>J Obstet Gynaecol</i> 2013; <b>33</b> :905–6. <a href="https://doi.org/10.3109/01443615.2013.834309">https://doi.org/10.3109/01443615.2013.834309</a>	Study design	Case study
Goepfert AR, Goldenberg RL, Mercer B, Iams J, Meis P, Moawad A, <i>et al.</i> 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. <i>Am J Obstet Gynecol</i> 2000; <b>183</b> :1480–3	Population	Asymptomatic population
Goldenberg RL, Iams JD, Mercer BM, Meis PJ, Moawad AH, Copper RL, <i>et al.</i> The preterm prediction study: the value of new vs standard risk factors in predicting early and all spontaneous preterm births. NICHD MFMU Network. <i>Am J Public Health</i> 1998; <b>88</b> :233–8	Population	Asymptomatic population
Goldenberg RL, Klebanoff M, Carey JC, Macpherson C, Leveno KJ, Moawad AH, <i>et al.</i> Vaginal fetal fibronectin measurements from 8 to 22 weeks' gestation and subsequent spontaneous preterm birth. <i>Am J Obstet Gynecol</i> 2000; <b>183</b> :469–75	Population	Asymptomatic population
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. <i>Obstet Gynecol</i> 1996; <b>87</b> :643–8	Population	Asymptomatic population

continued

TABLE 29 Hologic, Inc. (fFN) submitted citations (continued)

Citation	Reason for exclusion	Further detail
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. <i>J Perinat Med</i> 2017; <b>45</b> :85–9. <a href="https://doi.org/10.1515/jpm-2015-0270">https://doi.org/10.1515/jpm-2015-0270</a>	Outcome	Does not report test accuracy data
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. <i>Eur J Obstet Gynecol Reprod Biol</i> 2017; <b>208</b> :103–8	Population	Asymptomatic population
Hezelgrave NL, Shennan AH. Quantitative fetal fibronectin to predict spontaneous preterm birth: a review. <i>Womens Health</i> 2016; <b>12</b> :121–8. <a href="https://doi.org/10.2217/whe.15.74">https://doi.org/10.2217/whe.15.74</a>	Population	Asymptomatic population
Hezelgrave NL, Abbott DS, Radford SK, Seed PT, Girling JC, Filmer J, <i>et al.</i> Quantitative Fetal Fibronectin at 18 Weeks of Gestation to Predict Preterm Birth in Asymptomatic High-Risk Women. <i>Obstet Gynecol</i> 2016; <b>127</b> :255–63. <a href="https://doi.org/10.1097/AOG.0000000000001240">https://doi.org/10.1097/AOG.0000000000001240</a>	Population	Asymptomatic population
Jwala S, Tran TL, Terenna C, McGregor A, Andrel J, Leiby BE, <i>et al.</i> Evaluation of additive effect of quantitative fetal fibronectin to cervical length for prediction of spontaneous preterm birth among asymptomatic low-risk women. <i>Acta Obstet Gynecol Scand</i> 2016; <b>95</b> :948–55. <a href="https://doi.org/10.1111/aogs.12907">https://doi.org/10.1111/aogs.12907</a>	Population	Asymptomatic population
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. <i>Ultrasound Obstet Gynecol</i> 2016; <b>47</b> :210–16. <a href="https://doi.org/10.1002/uog.14894">https://doi.org/10.1002/uog.14894</a>	Comparator	Outcome at 14 days
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. <i>Ultrasound Obstet Gynecol</i> 2016; <b>47</b> :104–9. <a href="https://doi.org/10.1002/uog.14865">https://doi.org/10.1002/uog.14865</a>	Population	Asymptomatic population
Kuhrt K, Unwin C, Hezelgrave N, Seed P, Shennan A. Endocervical and high vaginal quantitative fetal fibronectin in predicting preterm birth. <i>J Matern Fetal Neonatal Med</i> 2014; <b>27</b> :1576–9. <a href="https://doi.org/10.3109/14767058.2013.870550">https://doi.org/10.3109/14767058.2013.870550</a>	Population	Asymptomatic population
Kurtzman J, Chandiramani M, Briley A, Poston L, Das A, Shennan A. Quantitative fetal fibronectin screening in asymptomatic high-risk patients and the spectrum of risk for recurrent preterm delivery. <i>Am J Obstet Gynecol</i> 2009; <b>200</b> :263.e1–6. <a href="https://doi.org/10.1016/j.ajog.2009.01.018">https://doi.org/10.1016/j.ajog.2009.01.018</a>	Population	Asymptomatic population
Lu GC, Goldenberg RL, Cliver SP, Kreaden US, Andrews WW. Vaginal fetal fibronectin levels and spontaneous preterm birth in symptomatic women. <i>Obstet Gynecol</i> 2001; <b>97</b> :225–8	Study design	ELISA test
McLaren JS, Hezelgrave NL, Ayubi H, Seed PT, Shennan AH. Prediction of spontaneous preterm birth using quantitative fetal fibronectin after recent sexual intercourse. <i>Am J Obstet Gynecol</i> 2015; <b>212</b> :89.e1–5. <a href="https://doi.org/10.1016/j.ajog.2014.06.055">https://doi.org/10.1016/j.ajog.2014.06.055</a>	Population	Asymptomatic population
Min J, Watson HA, Hezelgrave NL, Seed PT, Shennan AH. Ability of a preterm surveillance clinic to triage risk of preterm birth: a prospective cohort study. <i>Ultrasound Obstet Gynecol</i> 2016; <b>48</b> :38–42. <a href="https://doi.org/10.1002/uog.15925">https://doi.org/10.1002/uog.15925</a>	Population	Asymptomatic population
Ridout A, Carter J, Shennan A. Clinical utility of quantitative fetal fibronectin in preterm labour. <i>BJOG</i> 2016; <b>123</b> :1972. <a href="https://doi.org/10.1111/1471-0528.13850">https://doi.org/10.1111/1471-0528.13850</a>	Study design	Letter



TABLE 29 Hologic, Inc. (fFN) submitted citations (continued)

Citation	Reason for exclusion	Further detail
Ross GN, Ridout AE, Shennan AH. Optimal clinical risk prediction can be achieved by combining quantitative fetal fibronectin and cervical length, and avoiding thresholds. <i>Acta Obstet Gynecol Scand</i> 2016; <b>95</b> :956. <a href="https://doi.org/10.1111/aogs.12922">https://doi.org/10.1111/aogs.12922</a>	Study design	Letter
Schindhelm RK, Hoogenberg J, de Vos MT, Tegelaers FP. Analytical performance of quantitative fetal fibronectin assay. <i>Ultrasound Obstet Gynecol</i> 2016; <b>47</b> :127. <a href="https://doi.org/10.1002/uog.15824">https://doi.org/10.1002/uog.15824</a>	Study design	Letter
van der Krogt L, Hezelgrave NL, Seed PT, Shennan AH. Prediction of spontaneous preterm birth using fetal fibronectin in women with a low-lying placenta. <i>J Matern Fetal Neonatal Med</i> 2017; <b>30</b> :313–16. <a href="https://doi.org/10.3109/14767058.2016.1171837">https://doi.org/10.3109/14767058.2016.1171837</a>	Population	Asymptomatic population
Vandermolen BI, Hezelgrave NL, Smout EM, Abbott DS, Seed PT, Shennan AH. Quantitative fetal fibronectin and cervical length to predict preterm birth in asymptomatic women with previous cervical surgery. <i>Am J Obstet Gynecol</i> 2016; <b>215</b> :480.e1–480.e10. <a href="https://doi.org/10.1016/j.ajog.2016.05.020">https://doi.org/10.1016/j.ajog.2016.05.020</a>	Population	Asymptomatic population
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. <i>Ultrasound Obstet Gynecol</i> 2017; <b>50</b> :342–6	Abstract	Use of telephone application
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. <i>J Matern Fetal Neonatal Med</i> 2015; <b>28</b> :153–7. <a href="https://doi.org/10.3109/14767058.2014.909801">https://doi.org/10.3109/14767058.2014.909801</a>	Population	Asymptomatic population

## Parsagen Diagnostics Inc. (PartoSure) submitted citations

TABLE 30 Parsagen Diagnostics Inc. (PartoSure) submitted citations

Citation	Reason for exclusion	Further detail
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. <i>J Obstet Gynaecol Res</i> 2017; <b>43</b> :1263–9. <a href="https://doi.org/10.1111/jog.13366">https://doi.org/10.1111/jog.13366</a>	Included	
Echebiri NC, McDoom MM, Aalto M, Pullen J, Doyle NM. Placental alpha-microglobulin-1 and combined traditional diagnostic test: a cost-benefit analysis. <i>Am J Obstet Gynecol</i> 2014; <b>123</b> :3S-5	Population	Ruptured membranes
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. <i>Am J Obstet Gynecol</i> 2015; <b>212</b> :77.e1–10. <a href="https://doi.org/10.1016/j.ajog.2014.07.028">https://doi.org/10.1016/j.ajog.2014.07.028</a>	Population	Ruptured membranes
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. <i>J Matern Fetal Neonatal Med</i> 2016; <b>29</b> (Suppl. 1):283	Abstract	Not enough information
Hadzi-Lega M, Maier JT, Helmer H, Hellmeyer L, Markova AD, Poposka A. Comparison of PAMG-1 and pHGFBP-1 tests for the prediction of preterm delivery in patients with preterm labor. <i>Open J Obstet Gynecol</i> 2017; <b>7</b> :358–68	Included	

continued

TABLE 30 Parsagen Diagnostics Inc. (PartoSure) submitted citations (continued)

Citation	Reason for exclusion	Further detail
Heverhagen A. Placental alpha microglobulin-1 in combination with transvaginal ultrasound for prediction of preterm birth. <i>J Perinat Med</i> 2015; <b>43</b> (Suppl. 1):240	Abstract	Not enough information
Konoplyannikov A, Lysyuk I, Sokolyan A, Pipia N, Apresyan S, Karasova A, <i>et al.</i> PAMG-1 biomarker test (PARTOSURE) in combination with transvaginal ultrasound for improved assessment of spontaneous preterm birth in patients with threatened preterm labor. <i>J Matern Fetal Neonatal Med</i> 2016; <b>29</b> (Suppl. 1):278	Abstract	Not enough information
Lotfi G, Faraz S, Al Swalhee N, Nasir R, Somini S, Abdeldayem R, <i>et al.</i> Evaluation of PAMG-1 for the prediction of preterm birth in patients symptomatic of preterm labour. <i>J Perinat Med</i> 2015; <b>43</b> (Suppl. 1):250	Abstract	Not enough information
Lou YY, Ajay B. Is PartoSure effective in assessing preterm birth? <i>BJOG</i> 2016; <b>123</b> (Suppl. 2):89	Abstract	Not enough information
Ravi M, Beljorie M, El Masry K. 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. <i>Journal of Pediatric and Neonatal Individualized Medicine</i> 2017; <b>6</b> :ABS50	Abstract	Not enough information
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. <i>J Perinat Med</i> 2014; <b>42</b> :473–7. <a href="https://doi.org/10.1515/jpm-2013-0234">https://doi.org/10.1515/jpm-2013-0234</a>	Included	
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. <i>J Perinat Med</i> 2015; <b>43</b> :395–402. <a href="https://doi.org/10.1515/jpm-2014-0300">https://doi.org/10.1515/jpm-2014-0300</a>	Included	
Nikolova T, Uotila J, Nikolova N, Borisova VY, Bolotskikh VM. 16: Do PAMG-1 or pHlGFBP-1 biomarkers improve the prediction of imminent spontaneous preterm delivery in PTL symptomatic women with non-obvious cervical length (CL)? <i>Am J Obstet Gynecol</i> 2017; <b>216</b> :S11–12	Abstract	Not enough information
Melchor JC, Navas H, Marcos M, Iza A, de Diego M, Rando D, Burgos J. <i>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.</i> Conference: 1st World Congress on Maternal Fetal Neonatal Medicine. London, April 2017	Abstract	Not enough information
Van Holsbeke C, Dam K, Staelens A, Mesens T, Corremans A. Comparison of the fetal fibronectin (Rapid fFN) and placental alpha microglobulin-1 (PartoSure) tests for predicting imminent spontaneous preterm birth. <i>Ultrasound Obstet Gynecol</i> 2016; <b>48</b> :84	Abstract	Not enough information
Wing D, Haeri S, Silber A, Roth C, Echebiri N, Franco A, Pappas L, <i>et al.</i> PAMG-1 (PARTOSURE™) vs. fFN to Assess Risk of Preterm Delivery in Symptomatic Women. Conference: KU Medical Centre/UC Irvine Health Institute. PartoSure GA Analysis	Abstract	Not enough information

## Studies excluded at full-text review, with reasons

**TABLE 31** Studies excluded at full-text review, with reasons

Citation	Reason for exclusion
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. <i>J Perinat Med Conference: 12th World Congress of Perinatal Medicine 2015</i> ;43(no pagination)	Abstract
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. <i>Clinical Chemistry</i> 2008;54:A39–40	Abstract
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. <i>Reprod Sci</i> 2014;1:155A	Abstract
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. <i>J Matern Fetal Neonatal Med</i> 2016;29:283	Abstract
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. <i>Am J Obstet Gynecol</i> 2002;187(Suppl. 6):S80	Abstract
Grobman WA, Welshman EE, Calhoun EA, Ramsey PS. Fetal fibronectin results did not reduce medical resource use for women with preterm uterine contractions. <i>Evidence-based Obstetrics and Gynecology</i> 2005;7:118–9	Abstract
Hansen W, Lowe M, Zimmerman B. Effect of the fetal fibronectin assay on preterm labor management. <i>Am J Obstet Gynecol</i> 2001;185(Suppl. 6):S136. URL: <a href="http://onlinelibrary.wiley.com/doi/cochrane/central/articles/596/CN-00387596/frame.html">http://onlinelibrary.wiley.com/doi/cochrane/central/articles/596/CN-00387596/frame.html</a> (accessed September 2017)	Abstract
Heverhagen A. Placental alphamicroglobulin-1 in combination with transvaginal ultrasound for prediction of preterm birth. <i>J Perinat Med Conference: 12th World Congress of Perinatal Medicine 2015</i> ;43(no pagination)	Abstract
Heverhagen A, Baumann M, Raio L, Surbek D. Placental alpha-microglobulin-1 in combination with transvaginal ultrasound for prediction of preterm birth. <i>Am J Obstet Gynecol</i> 2015;212(Suppl. 1):S81	Abstract
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. <i>Reprod Sci</i> 2016;1:131A–2A	Abstract
Hillman-Cooper C, Ghag K, Dempsey A, Denbow M, Lopez Bernal A. Actim Partus-the first year at St. Michael's Hospital, Bristol. <i>Arch Dis Child Fetal Neonatal Ed</i> 2014;99:A158–9	Abstract
Holmgren C, Lacoursiere DY, Esplin MS. Clinical predictors of a false negative fetal fibronectin (FFN). <i>Am J Obstet Gynecol</i> 2007;197:S204-S	Abstract
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. <i>Am J Obstet Gynecol</i> 2007;197(Suppl. 6):S47	Abstract
Karunakaran B, Berry J, Parasuraman R. Can we raise the threshold for negative fetal fibronectin result? <i>Arch Dis Child Fetal Neonatal Ed</i> 2012;97:A89	Abstract
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. <i>J Matern Fetal Neonatal Med Conference: 25th European Congress of Perinatal Medicine, the Netherlands 2016</i> ;29:278. URL: <a href="http://onlinelibrary.wiley.com/doi/cochrane/central/articles/387/CN-01214387/frame.html">http://onlinelibrary.wiley.com/doi/cochrane/central/articles/387/CN-01214387/frame.html</a> (accessed September 2017)	Abstract

continued

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Kuhnert M. Individual management of preterm labour by using rapid fetal fibronectin test. <i>J Matern Fetal Neonatal Med</i> 2014; <b>27</b> :392–3	Abstract
Kung RWK, Northridge R, Nicoll AE. Fetal fibronectin (FFN) and iatrogenic pre-term birth. <i>Arch Dis Child Fetal Neonatal Ed</i> 2012; <b>97</b> :A80–1	Abstract
Lawin-O'Brien A, Jesner O, Biswas C. Evaluation of fetal fibronectin testing, subsequent management and outcomes in a London teaching hospital. <i>Arch Dis Child Fetal Neonatal Ed</i> 2014; <b>99</b> :A119	Abstract
Lega MH. Prediction of preterm delivery in patients in preterm labor. <i>J Perinat Med Conference: 12th World Congress of Perinatal Medicine</i> 2015; <b>43</b> (no pagination)	Abstract
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. <i>J Perinat Med Conference: 12th World Congress of Perinatal Medicine</i> 2015; <b>43</b> (no pagination)	Abstract
Lou YY, Ajay B. Is PartoSure effective in assessing preterm birth? <i>BJOG</i> 2016; <b>123</b> :89	Abstract
Mahavarker S, Osman MW, Ten-Hof J. Phosphorylated insulin-like growth factor binding protein-1 (ACTIM PARTUS) testing in management of threatened preterm labour. <i>Arch Dis Child Fetal Neonatal Ed</i> 2012; <b>97</b> :A99–100	Abstract
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. <i>J Matern Fetal Neonatal Med</i> 2016; <b>29</b> :56	Abstract
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. <i>J Matern Fetal Neonatal Med</i> 2016; <b>29</b> :56	Abstract
Nikolova T, Uotila J, Nikolova N, Borisova V, Bolotskikh V. Do PAMG-1 or pHGFBP-1 biomarkers improve the prediction of imminent spontaneous preterm delivery in PTL symptomatic women with non-obvious cervical length (CL)? <i>Am J Obstet Gynecol Conference: 37th annual meeting of the society for maternal-fetal medicine: the pregnancy meeting United States</i> 2017; <b>216</b> (Suppl. 1):S11–2. URL: <a href="http://onlinelibrary.wiley.com/doi/10.1002/ajog.156">http://onlinelibrary.wiley.com/doi/10.1002/ajog.156</a> (accessed September 2017)	Abstract
Northridge R, Liu H, Youssef R, Nicoll AE. Predicting pre-term birth using fetal fibronectin (FFN) in Ninewells Hospital, Dundee. <i>Arch Dis Child Fetal Neonatal Ed</i> 2011; <b>96</b> :Fa69	Abstract
Ponnusamy V, Farrer K. Impact of antenatal tests on acute in-utero transfers. <i>Arch Dis Child Fetal Neonatal Ed</i> 2012; <b>97</b> :A87	Abstract
Radford S, Costa F, Junior E, Sheehan P. Clinical application of quantitative fetal fibronectin for the prediction of preterm birth in symptomatic women. <i>BJOG: Conference: RCOG World Congress 2016, UK</i> 2016; <b>123</b> :158–9. URL: <a href="http://onlinelibrary.wiley.com/doi/10.1111/1471-0541.14790">http://onlinelibrary.wiley.com/doi/10.1111/1471-0541.14790</a> (accessed September 2017)	Abstract
Randhawa TS, Kwek K. Assessment of the predictive value of phosphorylated insulinlike growth factor binding protein-1 for preterm delivery. <i>J Matern Fetal Neonatal Med</i> 2010; <b>23</b> :294	Abstract
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. <i>BJOG</i> 2016; <b>123</b> :8	Abstract
Thandayathany V, Yassin MAJM, Omar MH, Ismail NAM, Tamil AM, Kampan NC. Fetal fibronectin rapid test versus phosphorylated insulin-like growth factor-1 (pHGFBP-1) as bedside test kits for prediction of preterm delivery in the clinical setting. <i>BJOG</i> 2012; <b>119</b> :5	Abstract
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. <i>Ultrasound Obstet Gynecol</i> 2016; <b>48</b> :84	Abstract
Sophie W, Tiphaine B, Celine C, Aurelie DB. Preterm labor: Reliability of the partosure test before and after digital and transvaginal ultrasound examinations. <i>J Perinat Med Conference: 12th World Congress of Perinatal Medicine</i> 2015; <b>43</b> (no pagination)	Abstract

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Bare T, Mitre A, Shahinaj R, Bakiri F. Fetal fibronectin and cericometry in patients at increased risk for premature birth. <i>J Perinat Med Conference: 12th World Congress of Perinatal Medicine</i> 2015; <b>43</b> (no pagination)	Abstract
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. <i>Am J Obstet Gynecol</i> 2009; <b>201</b> :S68–S9	Abstract
Berghella V, Hayes E, Visintine J, Baxter JK. Fetal fibronectin testing for reducing the risk of preterm birth. <i>Cochrane Database Syst Rev</i> 2008; <b>4</b> :CD006843. <a href="https://doi.org/10.1002/14651858.CD006843.pub2">https://doi.org/10.1002/14651858.CD006843.pub2</a>	Abstract
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. <i>Am J Obstet Gynecol</i> 1995; <b>172</b> :405	Abstract
Burrus DR, Ernest JM, Veille JC. Efficacy of four markers of preterm labor (PTL) at predicting latency (LAT) in a population receiving tocolytic therapy. <i>Am J Obstet Gynecol</i> 1995; <b>172</b> :405	Abstract
Goepfert AR, Network NM. The preterm prediction study: Quantitative fetal fibronectin (FFN) values and the prediction of spontaneous preterm birth (SPTB). <i>Am J Obstet Gynecol</i> 1997; <b>176</b> :S52	Abstract
Goldenberg RL. Vaginal fetal fibronectin (V-FFN) levels at 8–22 weeks and subsequent spontaneous preterm birth (SPB). <i>Am J Obstet Gynecol</i> 2000; <b>182</b> :S32	Abstract
Greenhagen JB, Van Wagoner J, Dudley D, Hunter C, Mitchell M, Casal D, <i>et al.</i> The value of fetal fibronectin as a predictor of preterm delivery in low risk women. <i>Am J Obstet Gynecol</i> 1996; <b>174</b> :471	Abstract
Grover CM. Use of the fetal fibronectin assay to avert unnecessary hospitalization and bedrest in patients with preterm labor symptoms: a cost analysis. <i>Prim Care Update Ob Gyns</i> 1998; <b>5</b> :178–9	Abstract
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 spontaneous preterm delivery in symptomatic women. <i>J Matern Fetal Neonatal Med</i> 2014; <b>27</b> :399	Abstract
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). <i>Am J Obstet Gynecol</i> 2005; <b>193</b> :S52-S	Abstract
Henrich W. Cervicometry and fibronectin role. <i>J Matern Fetal Neonatal Med</i> 2010; <b>23</b> :39	Abstract
Hillman-Cooper C, Denbow M, Lopez Bernal A. Identification and quantification of maternal plasma proteins that predict preterm labour. <i>BJOG</i> 2013; <b>120</b> :160–1	Abstract
Hillman-Cooper CS, Denbow M, Bernal AL. Identification and quantification of maternal plasma proteins that predict preterm labour. <i>Endocrine Reviews Conference: 95th Annual Meeting and Expo of the Endocrine Society, ENDO</i> 2013; <b>34</b> (Suppl. 1)	Abstract
Hong J-S, Yoon BH, Romero R, Shim S-S, Sohn Y-K, Park C-W, <i>et al.</i> Patients with a positive cervical fetal fibronectin followed by a negative test have a reduced likelihood of spontaneous preterm delivery. <i>Am J Obstet Gynecol</i> 2003; <b>189</b> (Suppl. 6):S173	Abstract
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. <i>Am J Obstet Gynecol</i> 2004; <b>191</b> :S111-S	Abstract
Joffe G. The use & cost benefit of fetal fibronectin testing in an integrated health care system. <i>Clinical Chemistry</i> 1999; <b>45</b> :S32–S3	Abstract
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? <i>BJOG</i> 2015; <b>122</b> :186	Abstract
Khizar SBK, Vallabu S. How to introduce a new service in clinical practice. <i>BJOG</i> 2013; <b>120</b> :96	Abstract
Kiefer D, Keeler S, Rust O, Demishev M, Muscat J, Bornstein E, <i>et al.</i> Is fetal fibronectin (FFN) a marker of intra-amniotic inflammation in patients with midtrimester short cervix? <i>Am J Obstet Gynecol</i> 2009; <b>201</b> :S60-S	Abstract
Knee A, Belisle E, Markenson G, Malshe A, Plevyak M, Bsaf F, <i>et al.</i> Do pregnancies complicated by preterm births with a negative fetal fibronectin screen have different characteristics compared to those that have positive fibronectin results? <i>Am J Obstet Gynecol</i> 2011; <b>204</b> :S190-S	Abstract

continued

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Knight B, Smith H, Tregarthen A, Lapointe J, Shennan A, Taylor MJO, <i>et al.</i> Quantitative fetal fibronectin as a predictor of time to delivery in women undergoing induction of labour: a pilot study. <i>BJOG</i> 2016; <b>123</b> :E5-E	Abstract
Kreaden U, Sheldon L, Lockhart T, Paulson M, Shorter S, Martinez R, <i>et al.</i> Natural distribution of fetal fibronectin (fFN) among symptomatic and asymptomatic pregnant women. <i>Clinical Chemistry</i> 1999; <b>45</b> :A86-7	Abstract
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. <i>Am J Obstet Gynecol</i> 1995; <b>172</b> :406	Abstract
Kuhr 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. <i>BJOG</i> 2013; <b>120</b> :14-15	Abstract
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). <i>Am J Obstet Gynecol</i> 2009; <b>1</b> :S227	Abstract
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. <i>Am J Obstet Gynecol</i> 2008; <b>199</b> :S212-S	Abstract
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. <i>Am J Obstet Gynecol</i> 2016; <b>214</b> :S441-2	Abstract
Luzzi V, Hankins K, Gronowski AM. Accuracy of the rapid fetal fibronectin TLI system in predicting preterm delivery. <i>Clin Chem</i> 2003; <b>49</b> :501-2	Abstract
Maharajan P, Elliot J, Thampi P. Efficacy of fibronectin and its impact on the use of atosiban-cost implication. <i>Arch Dis Child Fetal Neonatal Ed</i> 2012; <b>97</b> :A92	Abstract
Mancuso M, Biggio J, Owen J. Fetal fibronectin and latency period in women with arrested preterm labor. <i>Am J Obstet Gynecol</i> 2013; <b>208</b> :S221-S	Abstract
Mateus J, Pereira L, Baxter J, Berghella V, Cheku B, Tolosa JE. Fetal fibronectin compared to cervical dilation in clinical practice. <i>Am J Obstet Gynecol</i> 2003; <b>189</b> :S165-S	Abstract
McKenna DS, Chung K, Shubert P, Iams JD. The effect of digital examination on the expression of fetal fibronectin. <i>Am J Obstet Gynecol</i> 1998; <b>178</b> :S188	Abstract
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. <i>Am J Obstet Gynecol</i> 2014; <b>210</b> :S224-S	Abstract
Mogami H, Kondo E, Fujita K, Chigusa Y, Nishimura F, Ujita M, <i>et al.</i> Fetal fibronectin is a pathogenic factor of preterm PROM and preterm birth. <i>Placenta</i> 2013; <b>34</b> :A4-A	Abstract
Morrison JC, Naefi RWI, Botti JJ, Katz M, Belluomini JM. Uterine activity and fetal fibronectin are independent and combined predictors of spontaneous preterm birth. <i>Am J Obstet Gynecol</i> 1995; <b>172</b> :406	Abstract
Muller A, McCullough M, Obican S, Norton H, Carter J, Gonzalez-Quintero VH. 'false' positive fetal fibronectin and rates of adverse obstetrical outcomes. <i>Am J Obstet Gynecol</i> 2006; <b>195</b> :S233-S	Abstract
Muzaffar S, Behrens R, Barns F, Walker S. Quantitative fetal fibronectin and preterm delivery. <i>BJOG</i> 2013; <b>120</b> :595-6	Abstract
Myint H, Singh V, Roberts NJ. Introducing fetal fibronectin testing to a District General Hospital in women with symptoms of preterm labour. <i>Arch Dis Child Fetal Neonatal Ed</i> 2011; <b>96</b> :Fa83	Abstract
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. <i>BJOG</i> 2013; <b>120</b> :41-2	Abstract
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. <i>Am J Obstet Gynecol</i> 2005; <b>193</b> :S58-S	Abstract

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
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. <i>Am J Obstet Gynecol</i> 1995; <b>172</b> :406	Abstract
O'Byrne L, Daly S. Audit of fetal fibronectin use in a tertiary referral hospital in the triaging of threatened preterm labour. <i>BJOG</i> 2016; <b>123</b> :13–14	Abstract
Osorio M, Neiva R, Montes L, Silva J, Pinelo S, Pinho M, et al. <i>The Impact of Fetal Fibronectin Assay on Preterm Labor Management: A Randomized Controlled Trial</i> . In Puertas A, Montoya F, Romero J, Hurtado J, Manzanares S, editors. Granada: European Congress, Advances in Perinatal Medicine; 2010	Abstract
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. <i>J Matern Fetal Neonatal Med</i> 2010; <b>23</b> :304–5	Abstract
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. <i>Am J Obstet Gynecol</i> 1996; <b>174</b> :303	Abstract
Pilpel S, Markenson G, Hoffman D. A comparison of fetal fibronectin testing in predicting preterm birth in twin versus singleton gestations. <i>Am J Obstet Gynecol</i> 2002; <b>187</b> (Suppl. 6):S120	Abstract
Pradhan-Drews M, Drews F, Goddard R. Audit on preterm labour. <i>BJOG</i> 2013; <b>120</b> :431–2	Abstract
Radford S, Abbott D, Seed P, Kemp J, Shennan A. Quantitative fetal fibronectin for the prediction of preterm birth in symptomatic women. <i>Arch Dis Child Fetal Neonatal Ed</i> 2012; <b>97</b> :A73	Abstract
Radford S, Smith J, Shennan A, Wallace EM. Interhospital transfers of women in threatened preterm labour to Victorian level 6 hospitals. <i>J Paediatr Child Health</i> 2015; <b>51</b> :51	Abstract
Ray D, Dyson D. Fetal fibronectin as a predictor of preterm delivery in symptomatic patients. <i>Am J Obstet Gynecol</i> 2001; <b>184</b> :S37	Abstract
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. <i>Am J Obstet Gynecol</i> 2001; <b>185</b> (Suppl. 6):S128	Abstract
Revah A, Sue-A-Quan A, Hannah M. Fetal fibronectin as a predictor of preterm birth: A systematic review of the literature. <i>Am J Obstet Gynecol</i> 1997; <b>176</b> :S53	Abstract
Richards KM, Troeger KA, Petrilla A, Agatep B, Johnsrud M, Blackwell SC. Burden of preterm labor in the texas medicaid population. <i>Value in Health</i> 2016; <b>19</b> :A178	Abstract
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. <i>Am J Obstet Gynecol</i> 1997; <b>176</b> :S6	Abstract
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. <i>J Soc Gynecol Investig</i> 2004; <b>11</b> :204A–5A	Abstract
Rosenblatt KP, Nicolaidis 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. <i>Reprod Sci</i> 2013; <b>1</b> :80A	Abstract
Salari K, Treadwell M, Thompson K. Predictors of preterm birth: a comparison of cervical length, fetal fibronectin and fetal adrenal gland enlargement. <i>Am J Obstet Gynecol</i> 2015; <b>212</b> :S179–80	Abstract
Sasson I, Triche EW, Azpuru H, Romano M, Paidas M. Short cervix and negative fetal fibronectin: low risk of delivery within 2 weeks of screening. <i>Am J Obstet Gynecol</i> 2008; <b>199</b> :S49-S	Abstract
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. <i>Ultraschall in der Medizin, Supplement Conference: Ultraschall</i> 2013; <b>34</b> (no pagination)	Abstract
Simhan H. Comparison of self-collected vs speculum-collected measurement of quantitative fetal fibronectin. <i>Am J Obstet Gynecol</i> 2016; <b>214</b> :S310–1	Abstract

continued

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Stern V, Amabebe E, Anumba D. The association between cervicovaginal fetal fibronectin, the metabolic markers of the vaginal microbiome, and preterm birth. <i>BJOG</i> 2016; <b>123</b> :79–80	Abstract
Stern V, Anumba D. Screening for preterm birth with serial cervical length measurement and fetal fibronectin quantification – does rate of change predict early delivery? <i>BJOG</i> 2016; <b>123</b> :80	Abstract
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. <i>Reprod Sci</i> 2016; <b>23</b> :106A-A	Abstract
Stern V, Healey J, Brown B, Anumba D. Assessing the risk of spontaneous premature birth by electrical impedance spectroscopy of the cervix: The eclippx study. <i>Reprod Sci</i> 2015; <b>22</b> :331A	Abstract
Stern V, Healey J, Jones G, Dixon S, Walters S, Brown B, et al. The ECCLIPPx study: Electrical impedance prediction of preterm birth. <i>Arch Dis Child Fetal Neonatal Ed</i> 2014; <b>99</b> :A113–6	Abstract
Swamy G, Simhan H, Seglin H, Heine P. The clinical utility of fetal fibronectin. <i>Am J Obstet Gynecol</i> 2001; <b>185</b> (Suppl. 6):S136	Abstract
Swancoat S. 'Recent vaginal exam' no longer a contraindication to fetal fibronectin sampling: A prospective study. <i>Obstetrics and Gynecology</i> 2016; <b>127</b> :16S-S	Abstract
Thompson KG, Vaillancourt C, Moore S. Air medical transfer of patients in preterm labour: current treatments and tocolytics provided in Ontario. <i>Canadian Journal of Emergency Medicine</i> 2013; <b>15</b> :S6	Abstract
Turitz AL, Ananth C, Gyamfi-Bannerman C. A comparison of quantitative pre- and post-vaginal manipulation fetal fibronectin. <i>Am J Obstet Gynecol</i> 2017; <b>216</b> (Suppl. 1):S197	Abstract
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? <i>Am J Obstet Gynecol</i> 2014; <b>210</b> :S395-S	Abstract
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. <i>BJOG</i> 2016; <b>123</b> :85	Abstract
Vanner T, Rooney M, Blackburn S. Trends in In-utero transfer: Royal Wolverhampton Hospitals NHS Trust 2007 to 2010. <i>Arch Dis Child Fetal Neonatal Ed</i> 2011; <b>96</b> :Fa38	Abstract
Watson H, Carter J, Tribe RM, Shennan AH. The QUIPP app: A safe alternative to a treat-all strategy for threatened preterm labour. <i>BJOG</i> 2017; <b>124</b> :30–1	Abstract
Watson HA, Ridout A, Ross G, Shennan AH. Decision-making about preterm birth using the QUIPP app: A survey of women's experiences. <i>BJOG</i> 2017; <b>124</b> :132	Abstract
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. <i>Eur J Obstet Gynecol Reprod Biol</i> 2005; <b>121</b> :159–63	Comparator
Akercan F, Kazandi M, Sendag F, Cirpan T, Mgoyi L, Terek MC, Sagol S. Value of cervical phosphorylated insulinlike growth factor binding protein-1 in the prediction of preterm labor. <i>J Reprod Med</i> 2004; <b>49</b> :368–72	Comparator
Gibson S, Hezelgrave NL, Shennan AH. The impact of quantitative fetal fibronectin on the clinical management of women symptomatic of preterm labour. <i>Arch Dis Child Fetal Neonatal Ed</i> 2014; <b>99</b> :A151–5	Comparator
Kuhr 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. <i>Ultrasound Obstet Gynecol</i> 2016; <b>47</b> :210–16. <a href="https://doi.org/10.1002/uog.14894">https://doi.org/10.1002/uog.14894</a>	Comparator
Honest H, Bachmann LM, Gupta JK, Kleijnen J, Khan KS. Cervicovaginal fetal fibronectin test for predicting risk of spontaneous preterm birth. <i>ACOG Clinical Review</i> 2002; <b>7</b> :3–4	Design
ACOG Committee. Management of preterm labor. <i>Int J Gynaecol Obstet</i> 2003; <b>82</b> :127–35	Design
Burwick R, Zork N, Lee G, Ross M, Kjos S. Cervical assessment of cervical length compared to fetal fibronectin in the prediction of preterm delivery in women with threatened preterm labor. <i>J Matern Fetal Med</i> 2011; <b>24</b> :127–31. <a href="https://doi.org/10.3109/14767058.2010.529201">https://doi.org/10.3109/14767058.2010.529201</a> . [Erratum published in <i>J Matern Fetal Med</i> 2016; <b>29</b> :3212.]	Design



TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
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. <i>J Obstet Gynaecol</i> 2013; <b>33</b> :414–15. <a href="https://doi.org/10.3109/01443615.2013.772129">https://doi.org/10.3109/01443615.2013.772129</a>	Design
Adams TM, Kinzler WL, Chavez MR, Fazzari MJ, Vintzileos AM. Practice patterns in the timing of antenatal corticosteroids for fetal lung maturity. <i>J Matern Fetal Neonatal Med</i> 2015; <b>28</b> :1598–601. <a href="https://doi.org/10.3109/14767058.2014.962508">https://doi.org/10.3109/14767058.2014.962508</a>	Design
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. <i>Am J Obstet Gynecol</i> 2011; <b>204</b> (Suppl. 1):S224	Design
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. <i>Obstetrics and Gynecology</i> 2016; <b>128</b> :404	Design
Bernhardt J, Dorman K. Pre-term birth risk assessment tools. Exploring fetal fibronectin and cervical length for validating risk. <i>AWHONN Lifelines</i> 2004; <b>8</b> :38–44	Design
Brink S, Chetwynd J, Mannix M. A telling test for early birth . . . use of a vaginal smear to detect fetal fibronectin. <i>US News &amp; World Report</i> 1995; <b>119</b> :86	Design
Bruijn MMC, Hermans FJR, Vis JY, Wilms FF, Oudijk MA, Kwee A, <i>et al.</i> Which factors contribute to false-positive, false-negative, and invalid results in fetal fibronectin testing in women with symptoms of preterm labor? <i>Am J Perinatol</i> 2017; <b>34</b> :234–9	Design
Chandiramani M, Di Renzo GC, Gottschalk E, Helmer H, Henrich W, Hoesli I, <i>et al.</i> Fetal fibronectin as a predictor of spontaneous preterm birth: a European perspective. <i>J Matern Fetal Neonatal Med</i> 2011; <b>24</b> :330–6. <a href="https://doi.org/10.3109/14767058.2010.496879">https://doi.org/10.3109/14767058.2010.496879</a>	Design
Chien PFW, Khan KS, Ogston S, Owen P. How useful is cervico-vaginal fetal fibronectin in predicting preterm labour – Evidence from an overview. <i>J Obstet Gynecol</i> 1998; <b>18</b> :95	Design
Colombo DF. Predicting spontaneous preterm birth – Fetal fibronectin and ultrasonography help to rule out labour, not rule it in. <i>BMJ</i> 2002; <b>325</b> :289–90	Design
Crawshaw SE, Briley AL. Clinical. A new strategy to tackle preterm labour: PREMETS trial. <i>RCM Midwives Journal</i> 1999; <b>2</b> :122–4	Design
Daskalakis GJ, Papantoniou NE, Koutsodimas NB, Papapanagiotou A, Antsaklis AJ. Fetal fibronectin as a predictor of preterm birth. <i>J Obstet Gynaecol</i> 2000; <b>20</b> :347–53	Design
Decimoni TC, Sansone D, Etto H, Santos AM, Araujo G, Fonseca ES, <i>et al.</i> Systematic literature review and cost-effectiveness analysis of fetal fibronectin test for predicting preterm in Brazil. <i>Value Health</i> 2013; <b>16</b> :A333–4	Design
Deshpande S, Asselt A, Tomini F, Armstrong N, Allen A, Noake C, <i>et al.</i> Rapid fetal fibronectin (fFN) testing to predict pre-term birth in women with symptoms of premature labour: A systematic review and cost-effectiveness analysis. <i>Health Technol Assess</i> 2013; <b>17</b> (4). URL: <a href="http://onlinelibrary.wiley.com/doi/10.1002/hta.23012">http://onlinelibrary.wiley.com/doi/10.1002/hta.23012</a> (accessed September 2017)	Design
Elliott JP, Miller HS, Coleman S, Rhea D, Abril D, Hallbauer K, <i>et al.</i> A randomized multicenter study to determine the efficacy of activity restriction for preterm labor management in patients testing negative for fetal fibronectin. <i>J Perinatol</i> 2005; <b>25</b> :626–30	Design
Evantash E. Vaginal Fetal Fibronectin to Predict Spontaneous Preterm Birth. <i>JAMA</i> 2017; <b>318</b> :198–9. <a href="https://doi.org/10.1001/jama.2017.7133">https://doi.org/10.1001/jama.2017.7133</a>	Design
Faron G, Boulvain M, Irion O. Fetal fibronectin for prediction of preterm delivery: A meta-analysis. <i>Am J Obstet Gynecol</i> 1996; <b>174</b> :471	Design
Feinberg RF, Kliman HJ. Fetal fibronectin and preterm labor. <i>N Engl J Med</i> 1992; <b>326</b> :708. <a href="https://doi.org/10.1056/NEJM199203053261013">https://doi.org/10.1056/NEJM199203053261013</a>	Design
French L. Fetal fibronectin to predict preterm delivery. <i>J Fam Pract</i> 1998; <b>47</b> :250–1	Design

continued

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Fuks B, Petrunin DD, Zارايسكي EI, Boltovskaya MN, Nazimova SV, Starosvetskaya NA, <i>et al.</i> , inventors Devices and methods for detecting amniotic fluid in vaginal secretions patent US 08114610. 2012 Feb 14 2012	Design
Fuks B, Petrunin DD, Zارايسكي Eli, Boltovskaya MN, Nazimova SV, Starosvetskaya NA, <i>et al.</i> , inventors Devices and methods for detecting amniotic fluid in vaginal secretions patent US 09568479. 2017 Feb 14 2017	Design
Gebhardt S, Odendaal HJ. Fetal fibronectin in vaginal secretions – a predictor of preterm delivery? <i>S Afr Med J</i> 1995; <b>85</b> :188	Design
Goldenberg RL, Andrews WW, Hauth JC. Markers of preterm birth. <i>Prenat Neonatal Med</i> 1998; <b>3</b> :43–6.	Design
Hee L. Likelihood ratios for the prediction of preterm delivery with biomarkers. <i>Acta Obstet Gynecol Scand</i> 2011; <b>90</b> :1189–99. <a href="https://doi.org/10.1111/j.1600-0412.2011.01187.x">https://doi.org/10.1111/j.1600-0412.2011.01187.x</a>	Design
Herbst A, Nilsson C. Diagnosis of early preterm labour. <i>BJOG</i> 2006; <b>113</b> (Suppl. 3):60–7	Design
Heyborne KD. Fetal fibronectin testing in threatened preterm labor: time for more study! <i>Am J Obstet Gynecol</i> 2017; <b>217</b> :94	Design
Hezelgrave NL, Shennan AH, David AL. Rational testing tests to predict imminent delivery in threatened preterm labour. <i>BMJ</i> 2015; <b>350</b> :2183	Design
Jeavons W. Sterile speculum exams & fFN collection. <i>AWHONN Lifelines</i> 2005; <b>9</b> :236–40	Design
Kiss H, Ahner R, Hohlagschwandtner M, Leitich H, Husslein P. Fetal fibronectin as a predictor of term labor: a literature review. <i>Acta Obstet Gynecol Scand</i> 2000; <b>79</b> :3–7	Design
Krupa FG, Faltin D, Cecatti JG, Surita FG, Souza JP. Predictors of preterm birth. <i>Int J Gynaecol Obstet</i> 2006; <b>94</b> :5–11	Design
Kuin RA, Vis JY, Mol BW. Fetal fibronectin as a short-term predictor of preterm birth in symptomatic patients. <i>Obstet Gynecol</i> 2010; <b>115</b> :186–7. <a href="https://doi.org/10.1097/AOG.0b013e3181c885ed">https://doi.org/10.1097/AOG.0b013e3181c885ed</a>	Design
Lamont RF, Dragovic B. A randomised controlled trial of metronidazole for the prevention of preterm birth in women positive for cervicovaginal fetal fibronectin: the PREMETS Study by Shennan <i>et al.</i> <i>BJOG</i> 2006; <b>113</b> :850–1	Design
McCubbin K, Moore S, MacDonald R, Vaillancourt C. Medical Transfer of Patients in Preterm Labor: Treatments and Tocolytics. <i>Prehosp Emerg Care</i> 2015; <b>19</b> :103–9. <a href="https://doi.org/10.3109/10903127.2014.942475">https://doi.org/10.3109/10903127.2014.942475</a>	Design
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). <i>Health Technol Assess</i> 2004; <b>4</b> (4)	Design
Musaad SM, Melson LC, Boswell RD. Fetal fibronectin assay may reduce management cost of preterm labour: an interval analysis. <i>Pathology</i> 2006; <b>38</b> :473–4	Design
National Institute for Health Research. <i>PartoSure™ Time to Delivery Test in Suspected Premature Labour</i> . National Institute for Health Research; 2013	Design
O'Sullivan M. Fetal fibronectin as a tool to reduce preterm labour admissions. <i>Australian Midwifery News</i> 2006; <b>6</b> :14–5	Design
Owen P. Fetal fibronectin detection for prediction of preterm birth in low risk women. <i>Br J Obstet Gynaecol</i> 1995; <b>102</b> :1019	Design
Plaut M. Fetal fibronectin and the safety of outpatient management of preterm labor symptoms. <i>Am J Obstet Gynecol</i> 1996; <b>174</b> :797–8	Design
Peterson WE, Sprague AE, Reszel J, Walker M, Fell DB, Perkins SL, <i>et al.</i> Women's perspectives of the fetal fibronectin testing process: a qualitative descriptive study. <i>BMC Pregnancy &amp; Childbirth</i> 2014; <b>14</b> :190	Design
Plaut M. Fetal fibronectin and the safety of outpatient management of preterm labor symptoms. <i>Am J Obstet Gynecol</i> 1996; <b>174</b> :797–8	Design

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
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. <i>Evidence-Based Practice</i> 2008; <b>11</b> :11–12	Design
Rizzo G, Arduini D. <i>Obstetrical Markes for Very Premature Delivery</i> . In Bevilacqua G, editor. Florence: 8 <sup>th</sup> World Congress of Perinatal Medicine; 2007	Design
Rizzo G, Capponi A, Angelini E, Muscatello A. <i>Ultrasonographic and Biochemical Markers of Preterm Birth</i> . In Carrera JM, Cabero L, Baraibar R, editors. Barcelona: 5 <sup>th</sup> World Congress of Perinatal Medicine; 2007	Design
Ross GN, Ridout AE, Shennan AH. Optimal clinical risk prediction can be achieved by combining quantitative fetal fibronectin and cervical length, and avoiding thresholds. <i>Acta Obstet Gynecol Scand</i> 2016; <b>95</b> :956. <a href="https://doi.org/10.1111/aogs.12922">https://doi.org/10.1111/aogs.12922</a>	Design
Sanchez-Ramos L. The preterm labor index and fetal fibronectin for prediction of preterm delivery with intact membranes. <i>Obstet Gynecol</i> 2003; <b>102</b> :196	Design
Schindhelm RK, Hoogenberg J, de Vos MT, Tegelaers FP. Analytical performance of quantitative fetal fibronectin assay. <i>Ultrasound Obstet Gynecol</i> 2016; <b>47</b> :127. <a href="https://doi.org/10.1002/uog.15824">https://doi.org/10.1002/uog.15824</a>	Design
Siassakos D, O'Brien K, Draycott T. Healthcare evaluation of the use of atosiban and fibronectin for the management of pre-term labour. <i>J Obstet Gynaecol</i> 2009; <b>29</b> :507–11. <a href="https://doi.org/10.1080/01443610903003191">https://doi.org/10.1080/01443610903003191</a>	Design
Vis JY, Wilms FF, Oudijk MA, Porath MM, Scheepers HC, Bloemenkamp KW, 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). <i>BMC Pregnancy Childbirth</i> 2009; <b>9</b> :38. <a href="https://doi.org/10.1186/1471-2393-9-38">https://doi.org/10.1186/1471-2393-9-38</a>	Design
Vivanti AJ, Benachi A, Rouzier R. 'Individualized assessment of preterm birth risk using two modified prediction models' from M. Mailath-Pokorny and colleagues. <i>Eur J Obstet Gynecol Reprod Biol</i> 2015; <b>188</b> :136–7. <a href="https://doi.org/10.1016/j.ejogrb.2015.03.013">https://doi.org/10.1016/j.ejogrb.2015.03.013</a>	Design
White K. Rapid test for fetal fibronectin measures preterm risk for women. <i>Journal of Womens Health</i> 1998; <b>7</b> :1076	Design
Wiqvist N. Fetal fibronectin and preterm birth. <i>Acta Obstetricia Et Gynecologica Scandinavica</i> 1993; <b>72</b> :507–8	Design
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. <i>Am J Obstet Gynecol</i> 2012; <b>1</b> :S219–20	Design
Bastek JA, Sammel MD, Srinivas SK, McShea MA, Foreman MN, Elovitz MA, Metlay JP. Clinical prediction rules for preterm birth in patients presenting with preterm labor. <i>Obstet Gynecol</i> 2012; <b>119</b> :1119–28. <a href="https://doi.org/10.1097/AOG.0b013e31825503e5">https://doi.org/10.1097/AOG.0b013e31825503e5</a>	Design
Tan H, Wen SW, Chen XK, Demissie K, Walker M. Early prediction of preterm birth for singleton, twin, and triplet pregnancies. <i>Eur J Obstet Gynecol Reprod Biol</i> 2007; <b>131</b> :132–7	Design
Abdelazim IA, Al-Sherbeeney 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. <i>Acta Medica International</i> 2016; <b>3</b> :69–74	Design
Akhmetgaliev AR, Fatkullin IF, Munavirova AA, Fatkullin FI. Algorithm to identify the signs of threatened preterm labour. <i>Kazanskii Meditsinskii Zhurnal</i> 2017; <b>98</b> :132–6	Language
Benoist G. [Prediction of preterm delivery in symptomatic women (preterm labor).] <i>J Gynecol Obstet Biol Reprod</i> 2016; <b>45</b> :1346–63	Language
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.] <i>Minerva Ginecol</i> 2011; <b>63</b> :477–83	Language
Florjański J, Klósek A, Zalewski J, Tomiałowicz M, Fuchs T, Heimrath J, Pajak J. [Ultrasonographic assessment of cervical length and measurement of fetal fibronectin level in cervical secretion of pregnant women in prophylaxis of premature deliveries.] <i>Ginekol Pol</i> 2001; <b>72</b> :1139–43	Language

continued

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Hampel 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.] <i>Geburtshilfe Frauenheilkd</i> 1994; <b>54</b> :685–90. <a href="https://doi.org/10.1055/s-2007-1023624">https://doi.org/10.1055/s-2007-1023624</a>	Language
Hampel 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. <i>Arch Gynecol Obstet</i> 1993; <b>254</b> :1470–2	Language
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. <i>Geburtshilfe Und Frauenheilkunde</i> 2001; <b>61</b> :755–60	Language
Hincz P, Wilczyński J, Pawłowicz P, Borowski D, Krekora M, Szaflik K. [Fetal fibronectin as a predictor of preterm delivery in patients with preterm contractions and cervical changes.] <i>Ginekol Pol</i> 2000; <b>71</b> :728–32	Language
Lorenz-Eberhardt G, Weiss PAM, Haas J, Puerstner P, Adelwoehrer NE. Premature delivery and fetal fibronectin in the cervical and vaginal secretion. <i>Geburtshilfe und Frauenheilkunde</i> 1994; <b>54</b> :414–6	Language
Mansouri A, Tadjerouni A, El Rabiey G, Baube S, Garnier G, Tribalat S. [Is fetal fibronectin a valid test predictive of premature labor?.] <i>Contracept Fertil Sex</i> 1997; <b>25</b> :380–4	Language
Roubille M, Maillivain A, Biguet-Vernier B, Bon C, Golfier F, Pichot JC. [Fetal fibronectin as predictor of preterm birth.] <i>Ann Biol Clin</i> 1999; <b>57</b> :93–7	Language
Surbek D, Bösigler H, Pavic N, Huber P, Almendral AC, Holzgreve W. [Fetal fibronectin as a marker of prematurity in a high risk patient sample.] <i>Z Geburtshilfe Neonatol</i> 1997; <b>201</b> :15–20	Language
Tchobroutsky C. Fetal fibronectin in cervical and vaginal secretions as a predictor of preterm delivery. <i>Contraception Fertilite Sexualite</i> 1998; <b>26</b> :23–5	Language
Velasco JG, Gonzalez AG. Febronectin and prevention of preterm delivery. <i>Medicina Clinica</i> 1995; <b>105</b> :54–5	Language
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.] <i>Gynecol Obstet Fertil</i> 2015; <b>43</b> :640–5. <a href="https://doi.org/10.1016/j.gyobfe.2015.07.002">https://doi.org/10.1016/j.gyobfe.2015.07.002</a>	Language
Wilms FF, van Stralen G, Porath MM, Papatsonis DN, Oei SG, Mol BW, Scherjon S. [Predicating imminent preterm labour based on a determination of foetal fibronectin in a vaginal smear.] <i>Ned Tijdschr Geneeskde</i> 2009; <b>153</b> :B398	Language
Abbott DS, Radford SK, Seed PT, Tribe RM, Shennan AH. Evaluation of a quantitative fetal fibronectin test for spontaneous preterm birth in symptomatic women. <i>Am J Obstet Gynecol</i> 2013; <b>208</b> :122.e1–6. <a href="https://doi.org/10.1016/j.ajog.2012.10.890">https://doi.org/10.1016/j.ajog.2012.10.890</a>	Outcome
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. <i>Front Med</i> 2016; <b>3</b> :48. <a href="https://doi.org/10.3389/fmed.2016.00048">https://doi.org/10.3389/fmed.2016.00048</a>	Outcome
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. <i>Gynecol Obstet Invest</i> 2009; <b>68</b> :1–8. <a href="https://doi.org/10.1159/000209394">https://doi.org/10.1159/000209394</a>	Outcome
Ben-Haroush A, Poran E, Yogeve 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. <i>J Matern Fetal Neonatal Med</i> 2010; <b>23</b> :854–6. <a href="https://doi.org/10.3109/14767050903300977">https://doi.org/10.3109/14767050903300977</a>	Outcome
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. <i>Am J Obstet Gynecol</i> 2017; <b>216</b> :S241-S	Outcome
Bogavac M, Simin N, Ranisavljević M, Budisić L. The role of insulin-like growth factor in prediction and prevention of preterm delivery. <i>Vojnosanit Pregl</i> 2010; <b>67</b> :883–6	Outcome
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? <i>Eur J Obstet Gynecol Reprod Biol</i> 2009; <b>146</b> :180–3	Outcome

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
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. <i>BJOG</i> 2015; <b>122</b> :104	Outcome
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. <i>Arch Dis Child Fetal Neonatal Ed Conference: 16th Annual Conference of the British Maternal and Fetal Medicine Society Dublin Ireland Conference Publication</i> 2013; <b>98</b> (no pagination)	Outcome
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. <i>Arch Dis Child Fetal Neonatal Ed</i> 2011; <b>96</b> :Fa65–6	Outcome
Echebiri NC, McDoom MM, Aalto M, Pullen J, Doyle NM. Placental alpha-microglobulin-1 and combined traditional diagnostic test: a cost–benefit analysis. <i>Obstetrics &amp; Gynecology</i> 2014; <b>123</b> :3S–5	Outcome
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. <i>Am J Perinatol</i> 2005; <b>22</b> :305–9. <a href="https://doi.org/10.1055/s-2005-870895">https://doi.org/10.1055/s-2005-870895</a>	Outcome
Fiorini F, Isted A, Hezelgrave NL, Shennan AH. Prediction of spontaneous preterm birth with quantitative fetal fibronectin in patients with bulging fetal membranes. <i>BJOG</i> 2015; <b>122</b> :123	Outcome
Fiorini F, Isted A, Hezelgrave NL, Shennan AH. Quantitative fetal fibronectin predicts preterm birth in women with bulging fetal membranes. <i>Eur J Obstet Gynecol Reprod Biol</i> 2016; <b>203</b> :127–31. <a href="https://doi.org/10.1016/j.ejogrb.2016.05.046">https://doi.org/10.1016/j.ejogrb.2016.05.046</a>	Outcome
Gardner M, Rouse D, Joffe G. Cost–benefit analysis of fetal fibronectin screening. <i>Am J Obstet Gynecol</i> 1998; <b>178</b> :S121	Outcome
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. <i>Am J Obstet Gynecol</i> 1998; <b>178</b> :S121	Outcome
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. <i>Am J Obstet Gynecol</i> 2001; <b>184</b> :483–8	Outcome
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. <i>Am J Obstet Gynecol</i> 2002; <b>187</b> (Suppl. 6):S80	Outcome
Hadzi Lega M, Daneva Markova A, Girevski V. Correlation of cervical length and different biochemical markers in spontaneous preterm birth up to 7 days in symptomatic patients. <i>J Perinat Med Conference: 12th World Congress of Perinatal Medicine</i> 2015; <b>43</b> (no pagination)	Outcome
Hadzi Lega M, Daneva Markova A, Stefanovic M, Tanturovski M. Interleukin 6 and fetal fibronectin as a predictors of preterm delivery in symptomatic patients. <i>Bosnian journal of basic medical sciences/Udruzenje basicnih medicinskih znanosti = Association of Basic Medical Sciences</i> 2015; <b>15</b> :51–6	Outcome
Hadzi Lega M, Daneva Markova A, Tanturovski M. Interleukin 6 and fetal fibronectin in prediction of preterm delivery in symptomatic patients. <i>J Matern Fetal Neonatal Med</i> 2014; <b>27</b> :373	Outcome
Hadzi-Lega M, Markova AD, Stefanovic M, Tanturovski M. Combination of selected biochemical markers and cervical length in the prediction of impending preterm delivery in symptomatic patients. <i>Clin Exp Obstet Gynecol</i> 2016; <b>43</b> :154–60	Outcome
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. <i>J Perinat Med</i> 2015; <b>43</b> :545–51. <a href="https://doi.org/10.1515/jpm-2014-0275">https://doi.org/10.1515/jpm-2014-0275</a>	Outcome
Kekki M, Kurki T, Kärkkäinen T, Hiilesmaa V, Paavonen J, Rutanen EM. Insulin-like growth factor-binding protein-1 in cervical secretion as a predictor of preterm delivery. <i>Acta Obstet Gynecol Scand</i> 2001; <b>80</b> :546–51	Outcome
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. <i>Medicinski Glasnik</i> 2008; <b>5</b> :121–4	Outcome

continued

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Paternoster D, Riboni F, Vitulo A, Plebani M, Dell'Avanzo M, Battagliarin G, <i>et al.</i> Phosphorylated insulin-like growth factor binding protein-1 in cervical secretions and sonographic cervical length in the prediction of spontaneous preterm delivery. <i>Ultrasound Obstet Gynecol</i> 2009; <b>34</b> :437–40. <a href="https://doi.org/10.1002/uog.6428">https://doi.org/10.1002/uog.6428</a>	Outcome
Paternoster DM, Cardascia L, Narne S, Parise A, De Paoli M, Plebani M, <i>et al.</i> <i>Predictive Markers of Preterm Labour</i> . In Cosmi EV, editor. Alghero: 2 <sup>nd</sup> International Congress New Technologies in Reproductive Medicine, Neonatology and Gynecology; 1999	Outcome
Paternoster DM, Muresan D, Vitulo A, Serena A, Battagliarin G, Dell'avanzo M, Nicolini U. Cervical pHIGFBP-1 in the evaluation of the risk of preterm delivery. <i>Acta Obstet Gynecol Scand</i> 2007; <b>86</b> :151–5	Outcome
Rahkonen L, Unkila-Kallio L, Nuutila M, Sainio S, Saisto T, Rutanen EM, Paavonen J. Cervical length measurement and cervical phosphorylated insulin-like growth factor binding protein-1 testing in prediction of preterm birth in patients reporting uterine contractions. <i>Acta Obstet Gynecol Scand</i> 2009; <b>88</b> :901–8. <a href="https://doi.org/10.1080/00016340903104281">https://doi.org/10.1080/00016340903104281</a>	Outcome
Větr M, Kudela M, Prášilová J. Fetal fibronectin in patients at increased risk for premature birth. <i>Acta Univ Palacki Olomuc Fac Med</i> 1996; <b>140</b> :55–7	Outcome
Centra M, Coata G, Picchiassi E, Alfonsi L, Meniconi S, Bini V, <i>et al.</i> Evaluation of quantitative fFn test in predicting the risk of preterm birth. <i>J Perinat Med</i> 2017; <b>45</b> :91–8. <a href="https://doi.org/10.1515/jpm-2015-0414">https://doi.org/10.1515/jpm-2015-0414</a>	Outcome
Anonymous. fFN testing not always so accurate. <i>Contemporary OB/GYN</i> 2009; <b>54</b> :18	Population
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. <i>Reprod Sci</i> 2013; <b>20</b> :124A-A	Population
Abbott D, Hezelgrave N, Seed P, Bennett P, Chandiramani M, David A, <i>et al.</i> EQUIPP: Evaluation of fetal fibronectin with a novel bedside quantitative instrument for the prediction of preterm birth. <i>J Matern Fetal Neonatal Med</i> 2014; <b>27</b> :368–9	Population
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. <i>BJOG</i> 2012; <b>119</b> :e1-e	Population
Audibert F, Fortin S, Delvin E, Djemli A, Brunet S, Dubé J, Fraser WD. Contingent use of fetal fibronectin testing and cervical length measurement in women with preterm labour. <i>J Obstet Gynaecol Can</i> 2010; <b>32</b> :307–12	Population
Calvert B, Hezelgrave N, Seed P, Shennan A. Quantitative fetal fibronectin and birth outcome in women with previous cervical surgery and a short cervix. <i>BJOG</i> 2015; <b>122</b> :102–3	Population
Esplin M. 6: The use of serial cervical length and quantitative fetal fibronectin to identify nulliparous women at risk of subsequent spontaneous preterm birth. <i>Am J Obstet Gynecol</i> 2016; <b>214</b> :S4–5	Population
Esplin MS, Elovitz MA, Iams JD, Parker CB, Wapner RJ, Grobman WA, <i>et al.</i> Predictive accuracy of serial transvaginal cervical lengths and quantitative vaginal fetal fibronectin levels for spontaneous preterm birth among nulliparous women. <i>JAMA</i> 2017; <b>317</b> :1047–56. <a href="https://doi.org/10.1001/jama.2017.1373">https://doi.org/10.1001/jama.2017.1373</a>	Population
Goldenberg R, Iams J, Mercer M, Meis P, Mowad A, Copper R, <i>et al.</i> Fetal fibronectin and spontaneous preterm birth. <i>Am J Obstet Gynecol</i> 1995; <b>172</b> :254	Population
Goldenberg RL. The preterm prediction study: Patterns of cervicovaginal fetal fibronectin (FFN) as predictors of subsequent FFN and spontaneous preterm delivery (SPD). <i>Am J Obstet Gynecol</i> 1997; <b>176</b> :S53	Population
Goldenberg RL. The preterm prediction study: Sequential cervical length and fetal fibronectin testing for the prediction of spontaneous preterm birth (SPB). <i>Am J Obstet Gynecol</i> 1999; <b>180</b> :S27	Population
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. <i>Aust N Z J Obstet Gynaecol</i> 2017; <b>57</b> :405–11. <a href="https://doi.org/10.1111/ajo.12610">https://doi.org/10.1111/ajo.12610</a>	Population
Jeurgens-Borst AJ, Bekkers RL, Sporken JM, van den Berg PP. Use of insulin like growth factor binding protein-1 in the diagnosis of ruptured fetal membranes. <i>Eur J Obstet Gynecol Reprod Biol</i> 2002; <b>102</b> :11–14	Population

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Kalafat E, Yuce T, Tanju O, Koc A. Preterm premature rupture of membrane assessment via transperineal ultrasonography: a diagnostic accuracy study. <i>J Matern Fetal Neonatal Med</i> 2016; <b>29</b> :3690–4. <a href="https://doi.org/10.3109/14767058.2016.1140742">https://doi.org/10.3109/14767058.2016.1140742</a>	Population
Kwek K, Khi C, Ting HS, Yeo GS. Evaluation of a bedside test for phosphorylated insulin-like growth factor binding protein-1 in preterm labour. <i>Ann Acad Med Singap</i> 2004; <b>33</b> :780–3	Population
Mešić Đogić L, Mičić D, Omeragić F, Kovač R, Fazlagić S. IGFBP-1 marker of cervical ripening and predictor of preterm birth. <i>Med Glas</i> 2016; <b>13</b> :118–24. <a href="https://doi.org/10.17392/856-16">https://doi.org/10.17392/856-16</a>	Population
Paternoster D, Pignataro R, Grella P. <i>Insulin-like Growth Factor Protein Test (IGFBP-1) in Predicting Preterm Delivery</i> . Cosmi EV, editor. Rome: 2 <sup>nd</sup> World Congress on Labour and Delivery; May 1997	Population
Paternoster DM, Pignataro R, Stella A, Bertoldini M, Braccianti M. <i>New Biotechnologies in the Prediction of Preterm Birth</i> . Cosmi EV, editor. Fuggi: 16 <sup>th</sup> International Congress on the Fetus as a Patient; 2000	Population
Sheehan A, Brennecke S, Boland R, Smith J. 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. <i>J Pediatr Child Health</i> 2017; <b>53</b> :90	Test
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. <i>BJOG</i> 2012; <b>119</b> :e1–2	Test
Abenheim HA, Morin L, Benjamin A. Does availability of fetal fibronectin testing in the management of threatened preterm labour affect the utilization of hospital resources? <i>Obstet Gynaecol Can</i> 2005; <b>27</b> :689–94	Test
Ai F, Li GQ, Jiang J, Dong XD. Neutrophil elastase and fetal fibronectin levels as predictors of single-birth prematurity. <i>Exp Ther Med</i> 2015; <b>10</b> :665–70. <a href="https://doi.org/10.3892/etm.2015.2508">https://doi.org/10.3892/etm.2015.2508</a>	Test
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. <i>J Perinat Med</i> 2007; <b>35</b> :36–42. <a href="https://doi.org/10.1515/JPM.2007.005">https://doi.org/10.1515/JPM.2007.005</a>	Test
Allen KA, Kiwi R. Clinical utilization of a rapid fetal fibronectin assay. <i>Ann Clin Lab Sci</i> 2000; <b>30</b> :323–4	Test
Andersen HF. Use of fetal fibronectin in women at risk for preterm delivery. <i>Clin Obstet Gynecol</i> 2000; <b>43</b> :746–58	Test
Anwar A, Lindow SW, Greaves L, Hall S, Jha R. The use of fetal fibronectin in suspected pre-term labour. <i>J Obstet Gynaecol</i> 2014; <b>34</b> :45–7. <a href="https://doi.org/10.3109/01443615.2013.823923">https://doi.org/10.3109/01443615.2013.823923</a>	Test
Armson A, Dodds L, Dooley K, Howlett A, McPhee A, Scott H. Fetal fibronectin testing for suspected preterm labour in Nova Scotia. <i>Am J Obstet Gynecol</i> 2004; <b>191</b> :S115–S	Test
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. <i>Am J Obstet Gynecol</i> 2008; <b>199</b> :S235–S	Test
Bare T, Manoku N. Fetal fibronectin and cericometry in patients at increased risk for premature birth. <i>Acta Microbiologica Hellenica</i> 2015; <b>60</b> :186	Test
Bartnicki J, Casal D, Kreaden US, Saling E, Vetter K. Fetal fibronectin in vaginal specimens predicts preterm delivery and very-low-birth-weight infants. <i>Am J Obstet Gynecol</i> 1996; <b>174</b> :971–4	Test
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 (WMMUH): potential to reduce bed occupancy further. <i>BJOG</i> 2015; <b>122</b> :97	Test
Benson JE, Landy HJ, Ghidini A, Drassinower D, Poggi SH. Fetal fibronectin for evaluation of preterm labor in the setting of cervical cerclage. <i>J Matern Fetal Neonatal Med</i> 2012; <b>25</b> :2330–2. <a href="https://doi.org/10.3109/14767058.2012.695820">https://doi.org/10.3109/14767058.2012.695820</a>	Test
Berghella V, Hayes E, Visintine J, Baxter JK. Fetal fibronectin testing for reducing the risk of preterm birth. <i>Cochrane Database Syst Rev</i> 2008; <b>4</b> :CD006843. <a href="https://doi.org/10.1002/14651858.CD006843.pub2">https://doi.org/10.1002/14651858.CD006843.pub2</a>	Test

continued

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
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. <i>Am J Obstet Gynecol</i> 2016; <b>215</b> :431–8. <a href="https://doi.org/10.1016/j.ajog.2016.04.038">https://doi.org/10.1016/j.ajog.2016.04.038</a>	Test
Bolt L, Chandiramani M, De Greeff A, Shennan A. Does fetal fibronectin testing change patient management in women at risk of pre-term labour? <i>Int J Gynecol Obstet</i> 2009; <b>107</b> :S527–8	Test
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. <i>Am J Obstet Gynecol</i> 2014; <b>210</b> :54.e1–54.e10. <a href="https://doi.org/10.1016/j.ajog.2013.09.004">https://doi.org/10.1016/j.ajog.2013.09.004</a>	Test
Bruijn M, Vis J, Wilms F, Oudijk M, Kwee A, Porath M, <i>et al.</i> The contribution of vaginal examination to risk stratification of women with signs of preterm labor before 34 weeks gestation: the APOSTEL1-cohort. <i>Am J Obstet Gynecol</i> 2014; <b>210</b> (Suppl. 1):S362–3	Test
Burrus D, Adair D, Aptel K, Ernest J, Frye A, Veille J. The effect of digital cervical exam on measured levels of fetal fibronectin. <i>Am J Obstet Gynecol</i> 1996; <b>174</b> :352	Test
Burrus DR, Ernest JM, Veille JC. Fetal fibronectin, interleukin-6, and C-reactive protein are useful in establishing prognostic subcategories of idiopathic preterm labor. <i>Am J Obstet Gynecol</i> 1995; <b>173</b> :1258–62	Test
Burwick RM, Zork NM, Lee GT, Ross MG, Kjos SL. Cervical assessment of cervical length compared to fetal fibronectin in the prediction of preterm delivery in women with threatened preterm labor. <i>J Matern Fetal Neonatal Med</i> 2011; <b>24</b> :127–31. <a href="https://doi.org/10.3109/14767058.2010.529201">https://doi.org/10.3109/14767058.2010.529201</a>	Test
Chandiramani M, Das A, Bolt LA, de Greeff A, Filmer J, Shennan A. Assessment of fetal fibronectin testing to improve preterm birth management. <i>Reprod Sci</i> 2014; <b>21</b> :151A–2A	Test
Cheung KW, Ngu SF, Lee CP. Fetal fibronectin test on Chinese women with symptoms of preterm labour: a pilot study. <i>Hong Kong Med J</i> 2013; <b>19</b> :424–8. <a href="https://doi.org/10.12809/hkmj133861">https://doi.org/10.12809/hkmj133861</a>	Test
Chien PF, Khan KS, Ogston S, Owen P. The diagnostic accuracy of cervico-vaginal fetal fibronectin in predicting preterm delivery: an overview (Structured abstract). <i>Br J Obstet Gynaecol</i> 1997; <b>104</b> :436–44. URL: <a href="http://onlinelibrary.wiley.com/doi/10.1002/obg.100650/frame.html">http://onlinelibrary.wiley.com/doi/10.1002/obg.100650/frame.html</a> (accessed September 2017)	Test
Clock C, Kurtzman J. Positive fetal fibronectin status is associated with dynamic cervical shortening in patients with normal baseline cervical length and symptomatic contractions. <i>Am J Obstet Gynecol</i> 2007; <b>197</b> :S199-S	Test
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. <i>Eur J Obstet Gynecol Reprod Biol</i> 2001; <b>95</b> :154–8	Test
Cox S, Little B, Dax J, Leveno K. Fetal fibronectin and preterm delivery. <i>Am J Obstet Gynecol</i> 1996; <b>174</b> :306	Test
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. <i>Am J Obstet Gynecol</i> 2013; <b>208</b> :233.e1–6. <a href="https://doi.org/10.1016/j.ajog.2012.12.015">https://doi.org/10.1016/j.ajog.2012.12.015</a>	Test
Deshpande SN, van Asselt AD, Tomini F, Armstrong N, Allen A, Noake C, <i>et al.</i> Rapid fetal fibronectin testing to predict preterm birth in women with symptoms of premature labour: a systematic review and cost analysis. <i>Health Technol Assess</i> 2013; <b>17</b> (40). <a href="https://doi.org/10.3310/hta17400">https://doi.org/10.3310/hta17400</a>	Test
Díaz 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. <i>J Matern Fetal Neonatal Med</i> 2009; <b>22</b> :89–93. <a href="https://doi.org/10.1080/14767050802464551">https://doi.org/10.1080/14767050802464551</a>	Test
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. <i>Arch Gynecol Obstet</i> 2011; <b>284</b> :559–65. <a href="https://doi.org/10.1007/s00404-010-1712-x">https://doi.org/10.1007/s00404-010-1712-x</a>	Test



TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
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. <i>Am J Perinatol</i> 2016; <b>33</b> :208–13. <a href="https://doi.org/10.1055/s-0035-1563710">https://doi.org/10.1055/s-0035-1563710</a>	Test
Elliot J, Maharajan P, Thampi P. Test that predicts premature labour: efficacy of foetal fibronectin test and its cost implications. <i>Arch Dis Child Fetal Neonatal Ed</i> 2012; <b>97</b> :A19–20	Test
Farag AH, Mohammed MM, Ellaithy MI, Salama HA. Blind vaginal fetal fibronectin swab for prediction of preterm birth. <i>J Obstet Gynaecol Res</i> 2015; <b>41</b> :1009–17. <a href="https://doi.org/10.1111/jog.12666">https://doi.org/10.1111/jog.12666</a>	Test
Fell DB, Sprague AE, Grimshaw JM, Yasseen AS, Coyle D, Dunn SI, <i>et al.</i> Evaluation of the impact of fetal fibronectin test implementation on hospital admissions for preterm labour in Ontario: a multiple baseline time-series design. <i>BJOG</i> 2014; <b>121</b> :438–46. <a href="https://doi.org/10.1111/1471-0528.12511">https://doi.org/10.1111/1471-0528.12511</a>	Test
Foong E, Chye Chuah S. Clinical application of quantitative fetal fibronectin testing in the management of preterm labour at a regional hospital. <i>Aust N Z J Obstet Gynaecol</i> 2016; <b>56</b> :37–8	Test
Foxman EF, Jarolim P. Use of the fetal fibronectin test in decisions to admit to hospital for preterm labor. <i>Clin Chem</i> 2004; <b>50</b> :663–5. <a href="https://doi.org/10.1373/clinchem.2003.028720">https://doi.org/10.1373/clinchem.2003.028720</a>	Test
Franco R, Shaw K, Williams C, Hickok D. Risk of preterm delivery with a positive fetal fibronectin test result and a normal cervical length. <i>Obstet Gynecol</i> 2005; <b>105</b> :114S–S	Test
Fricke E. A quality analysis of the evaluation of preterm labor. <i>Am J Obstet Gynecol</i> 2015; <b>1</b> :S66	Test
Gaikwad VA, Rajput UG, Puri MS. Cervico vaginal fluid fibronectin level in assessing risk of preterm labour. <i>Res J Pharm Biol Chem Sci</i> 2014; <b>5</b> :156–62	Test
Gao L, Zhang JP, Chen H, Guo ZJ, Chen LB, Tan JP, <i>et al.</i> Fetal fibronectin detection for preterm birth prediction. <i>Genet Mol Res</i> 2014; <b>13</b> :1323–8. <a href="https://doi.org/10.4238/2014.February.28.4">https://doi.org/10.4238/2014.February.28.4</a>	Test
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. <i>Am J Obstet Gynecol</i> 2005; <b>193</b> :S52–S	Test
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. <i>Am J Obstet Gynecol</i> 2000; <b>182</b> :439–42	Test
Goepfert AR, Andrews WW. The preterm prediction study: association between cervical interleukin-6 (IL-6), fetal fibronectin (FFN), and spontaneous preterm birth (SPTB). <i>Am J Obstet Gynecol</i> 1997; <b>176</b> :S6	Test
Rossel Goffeng A, Holst E, Milsom I, Lindstedt G, Lundberg PA, Andersch B. Factors influencing vaginal fetal fibronectin concentrations during pregnancy. <i>Gynecol Obstet Invest</i> 1999; <b>47</b> :83–8	Test
Goffinet F, Kayem G, Maillard F, Trébédén H, Cabrol D, Weill B, Batteux F. 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. <i>Eur J Obstet Gynecol Reprod Biol</i> 2005; <b>123</b> :167–73	Test
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. <i>J Perinat Med</i> 2017; <b>45</b> :85–9. <a href="https://doi.org/10.1515/jpm-2015-0270">https://doi.org/10.1515/jpm-2015-0270</a>	Test
Grobman WA, Welshman EE, Calhoun EA. Does fetal fibronectin use in the diagnosis of preterm labor affect physician behavior and health care costs? A randomized trial. <i>Am J Obstet Gynecol</i> 2004; <b>191</b> :235–40. <a href="https://doi.org/10.1016/j.ajog.2003.11.034">https://doi.org/10.1016/j.ajog.2003.11.034</a>	Test
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. <i>Aust N Z J Obstet Gynaecol</i> 2006; <b>46</b> :440–5	Test
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). <i>Health Technol Assess</i> 2008(4). URL: <a href="http://onlinelibrary.wiley.com/doi/10.1002/hta.32008100049/frame.html">http://onlinelibrary.wiley.com/doi/10.1002/hta.32008100049/frame.html</a> (accessed September 2017)	Test

continued

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Helmer H, Brunbauer M, Rohrmeister K. Exploring the role of Tractocile in everyday clinical practice. <i>BJOG</i> 2003; <b>110</b> (Suppl. 20):113–15	Test
Heng YJ, Pennell CE, Chua HN, Perkins JE, Lye SJ. Whole blood gene expression profile associated with spontaneous preterm birth in women with threatened preterm labor. <i>PLOS ONE</i> 2014; <b>9</b> :e96901. <a href="https://doi.org/10.1371/journal.pone.0096901">https://doi.org/10.1371/journal.pone.0096901</a>	Test
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 34 weeks and not delivering within 7 days. <i>Acta Obstetrica Et Gynecologica Scandinavica</i> 2015; <b>94</b> :715–21	Test
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. <i>Am J Obstet Gynecol</i> 2000; <b>182</b> :S44	Test
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. <i>Acta Obstet Gynecol Scand</i> 2002; <b>81</b> :58–63	Test
Honest H, Forbes CA, Durée KH, Norman G, Duffy SB, Tsourapas A, et al. Screening to prevent spontaneous preterm birth: systematic reviews of accuracy and effectiveness literature with economic modelling. <i>Health Technol Assess</i> 2009; <b>13</b> (43). <a href="https://doi.org/10.3310/hta13430">https://doi.org/10.3310/hta13430</a>	Test
Iams JD, Casal D, McGregor JA, Goodwin TM, Kreaden US, Lowensohn R, Lockitch G. Fetal fibronectin improves the accuracy of diagnosis of preterm labor. <i>Am J Obstet Gynecol</i> 1995; <b>173</b> :141–5	Test
Ibrahim MI, Sherif A, El-Kady M, Ellaithy M, Hussein A, Kamal M, El-Din NN. Can three-dimensional ultrasound measurement of fetal adrenal gland enlargement predict preterm birth? <i>Arch Gynecol Obstet</i> 2015; <b>292</b> :569–78. <a href="https://doi.org/10.1007/s00404-015-3668-3">https://doi.org/10.1007/s00404-015-3668-3</a>	Test
Incerti M, Ghidini A, Korker V, Pezzullo JC. Performance of cervicovaginal fetal fibronectin in a community hospital setting. <i>Arch Gynecol Obstet</i> 2007; <b>275</b> :347–51. <a href="https://doi.org/10.1007/s00404-006-0267-3">https://doi.org/10.1007/s00404-006-0267-3</a>	Test
Joffe G, Jacques D, Burton B, Bemis-Heys R, Shelburne P, Lovelace. Impact of fetal fibronectin assay on admissions for preterm labor. <i>Am J Obstet Gynecol</i> . 1998; <b>178</b> (1 PART 2):S11	Test
Joffe GM, Jacques D, Bemis-Heys R, Burton R, Skram B, Shelburne P. Impact of the fetal fibronectin assay on admissions for preterm labor. <i>Am J Obstet Gynecol</i> 1999; <b>180</b> :581–6	Test
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. <i>Comptes Rendus De L Academie Bulgare Des Sciences</i> 2016; <b>69</b> :661–6	Test
Keeler S, Roman A, Coletta J, Kiefer D, Rust O. Fetal fibronectin testing prior to ultrasound-indicated CERC1 age: can it predict procedure success? <i>Am J Obstet Gynecol</i> 2007; <b>197</b> :S199-S	Test
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? <i>Am J Obstet Gynecol</i> 2009; <b>200</b> :158.e1–6	Test
Langer B, Boudier E, Schlaeder G. Cervico-vaginal fetal fibronectin: predictive value during false labor. <i>Acta Obstet Gynecol Scand</i> 1997; <b>76</b> :218–21	Test
Lee G, Burwick R, Zork N, Kjos S. Does the use of fetal fibronectin in an algorithm for preterm labor reduce triage evaluation times? <i>J Matern Fetal Neonatal Med: the Official Journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians</i> 2013; <b>26</b> :706–9. URL: <a href="http://onlinelibrary.wiley.com/doi/10.1002/jmfm.1214">http://onlinelibrary.wiley.com/doi/10.1002/jmfm.1214</a> (accessed September 2017)	Test
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. <i>J Matern Fetal Neonatal Med</i> 2012; <b>25</b> :1690–8. <a href="https://doi.org/10.3109/14767058.2012.657279">https://doi.org/10.3109/14767058.2012.657279</a>	Test
Leitich H, Egarter C, Kaider A, Hohlagschwandtner M, Berghammer P, Husslein P. Cervicovaginal fetal fibronectin as a marker for preterm delivery: a meta-analysis. <i>Am J Obstet Gynecol</i> 1999; <b>180</b> :1169–76	Test

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Lopez RL, Francis JA, Garite TJ, Dubyak JM. Fetal fibronectin detection as a predictor of preterm birth in actual clinical practice. <i>Am J Obstet Gynecol</i> 2000; <b>182</b> :1103–6	Test
Lowe MP, Zimmerman B, Hansen W. Prospective randomized controlled trial of fetal fibronectin on preterm labor management in a tertiary care center. <i>Am J Obstet Gynecol</i> 2004; <b>190</b> :358–62. <a href="https://doi.org/10.1016/j.ajog.2003.08.041">https://doi.org/10.1016/j.ajog.2003.08.041</a>	Test
Lu GC, Goldenberg RL, Cliver SP, Kreaden US, Andrews WW. Vaginal fetal fibronectin levels and spontaneous preterm birth in symptomatic women. <i>Obstet Gynecol</i> 2001; <b>97</b> :225–8	Test
MacDonald WA, Bender M, Saxton A. Use of fetal fibronectin in the management of preterm labour in Nunavut. <i>Alaska Med</i> 2007; <b>49</b> (Suppl. 2):215–17	Test
Malak TM, Sizmur F, Bell SC, Taylor DJ. Fetal fibronectin in cervicovaginal secretions as a predictor of preterm birth. <i>Br J Obstet Gynaecol</i> 1996; <b>103</b> :648–53	Test
Marvin KW, Keelan JA, Coleman MA, McCowan LM, Zhou RL, Mitchell MD. Interleukin-1 (ICAM-1) in cervicovaginal fluid of women presenting with preterm labor: predictive value for preterm delivery. <i>Am J Reprod Immunol</i> 2000; <b>43</b> :264–71	Test
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. <i>Am J Perinatol</i> 2007; <b>24</b> :381–5. <a href="https://doi.org/10.1055/s-2007-981849">https://doi.org/10.1055/s-2007-981849</a>	Test
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. <i>Am J Obstet Gynecol</i> 1993; <b>168</b> :538–42	Test
Most O, Langer O, Kerner R, David GB, Calderon I. Can myometrial electrical activity identify patients in preterm labor? <i>Am J Obstet Gynecol</i> 2008; <b>199</b> :378.e1–6. <a href="https://doi.org/10.1016/j.ajog.2008.08.003">https://doi.org/10.1016/j.ajog.2008.08.003</a>	Test
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. <i>Pathology</i> 2005; <b>37</b> :226–30	Test
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. <i>Am J Obstet Gynecol</i> 2007; <b>197</b> :426.e1–7	Test
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. <i>Am J Obstet Gynecol</i> 2006; <b>195</b> (Suppl. 1):S67. URL: <a href="http://onlinelibrary.wiley.com/doi/10.1002/ajog.10777">http://onlinelibrary.wiley.com/doi/10.1002/ajog.10777</a> (accessed September 2017)	Test
Nguyen T. The cost-effectiveness of fetal fibronectin testing in suspected preterm labor: a randomized trial. <i>Obstet Gynecol</i> 2002; <b>99</b> (Suppl. 4):97s. URL: <a href="http://onlinelibrary.wiley.com/doi/10.1002/ajog.10777">http://onlinelibrary.wiley.com/doi/10.1002/ajog.10777</a> (accessed September 2017)	Test
Owen P, Scott A. Can fetal fibronectin testing improve the management of preterm labour? <i>Clin Exp Obstet Gynecol</i> 1997; <b>24</b> :19–22	Test
Parisaei M, Currie J, O’Gorman N, Morris S, David AL. Implementation of foetal fibronectin testing: admissions, maternal interventions and costs at 1 year. <i>J Obstet Gynaecol</i> 2016; <b>36</b> :888–92	Test
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. <i>Aust N Z J Obstet Gynaecol</i> 1995; <b>35</b> :257–61	Test
Parry E, Singh TG, Dow D, Noovao F. Improved management in threatened preterm labour with rapid fetal fibronectin testing. <i>Aust N Z J Obstet Gynaecol</i> 2006; <b>46</b> :240–1	Test
Plaut MM, Smith W, Kennedy K. Fetal fibronectin: the impact of a rapid test on the treatment of women with preterm labor symptoms. <i>Am J Obstet Gynecol</i> 2003; <b>188</b> :1588–93	Test
Revah A, Hannah ME, Sue-A-Quan AK. Fetal fibronectin as a predictor of preterm birth: an overview. <i>Am J Perinatol</i> 1998; <b>15</b> :613–21. <a href="https://doi.org/10.1055/s-2007-994079">https://doi.org/10.1055/s-2007-994079</a>	Test

continued

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Rinehart BK, Terrone DA, Isler CM, Barrilleaux PS, Bufkin L, Morrison JC. Pregnancy outcome in women with preterm labor symptoms without cervical change. <i>Am J Obstet Gynecol</i> 2001; <b>184</b> :1004–7	Test
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. <i>Am J Obstet Gynecol</i> 1996; <b>175</b> :1146–51	Test
Roman AS, Koklanaris N, Paidas MJ, Mulholland J, Levitz M, Rebarber A. 'Blind' vaginal fetal fibronectin as a predictor of spontaneous preterm delivery. <i>Obstet Gynecol</i> 2005; <b>105</b> :285–9	Test
Roos C, Schuit E, Scheepers HC, Bloemenkamp KW, Bolte AC, Duvekot HJ, et al. Predictive factors for delivery within 7 days after successful 48-hour treatment of threatened preterm labor. <i>AJP Rep</i> 2015; <b>5</b> :e141–9. <a href="https://doi.org/10.1055/s-0035-1552930">https://doi.org/10.1055/s-0035-1552930</a>	Test
Roos C, Vis JY, Scheepers HC, Bloemenkamp KW, Duvekot HJ, van Eyck J, et al. Fetal fibronectin status and cervical length in women with threatened preterm labor and the effectiveness of maintenance tocolysis. <i>J Matern Fetal Neonatal Med</i> 2016; <b>29</b> :1556–61. <a href="https://doi.org/10.3109/14767058.2015.1053863">https://doi.org/10.3109/14767058.2015.1053863</a>	Test
Rose CH, McWeeney DT, Brost BC, Davies NP, Watson WJ. Cost-effective standardization of preterm labor evaluation. <i>Am J Obstet Gynecol</i> 2010; <b>203</b> :250.e1–5. <a href="https://doi.org/10.1016/j.ajog.2010.06.037">https://doi.org/10.1016/j.ajog.2010.06.037</a>	Test
Rossel Goffeng A, Holst E, Milsom I, Lindstedt G, Lundberg PA, Andersch B. Factors influencing vaginal fetal fibronectin concentrations during pregnancy. <i>Gynecol Obstet Invest</i> 1999; <b>47</b> :83–8	Test
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. <i>Am J Obstet Gynecol</i> 1997; <b>176</b> :196–9	Test
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. <i>Obstet Gynecol</i> 2003; <b>101</b> :123–8	Test
Salcedo A, Kurtzman J. Digital cervical exam characteristics stratify the risk of preterm birth based on fetal fibronectin status. <i>Am J Obstet Gynecol</i> 2013; <b>208</b> :S203–5	Test
Senden IP, Owen P. Comparison of cervical assessment, fetal fibronectin and fetal breathing in the diagnosis of preterm labour. <i>Clin Exp Obstet Gynecol</i> 1996; <b>23</b> :5–9	Test
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. <i>J Obstet Gynaecol Can</i> 2006; <b>28</b> :206–13	Test
Stevens AO, Chauhan SP, Magann EF, Martin RW, Bofill JA, Cushman JL, Morrison JC. Fetal fibronectin and bacterial vaginosis are associated with preterm birth in women who are symptomatic for preterm labor. <i>Am J Obstet Gynecol</i> 2004; <b>190</b> :1582–7. <a href="https://doi.org/10.1016/j.ajog.2004.03.059">https://doi.org/10.1016/j.ajog.2004.03.059</a>	Test
Stuart RW, Jones RO, Underwood F, Kryka R, Bodor GS. The utility of fetal fibronectin in preterm labor decision making. <i>Am J Clin Path</i> 2007; <b>128</b> :513–4	Test
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. <i>J Obstet Gynaecol</i> 2013; <b>33</b> :566–8. <a href="https://doi.org/10.3109/01443615.2013.807786">https://doi.org/10.3109/01443615.2013.807786</a>	Test
Sullivan A, Hueppchen NA, Satin AJ. Cost effectiveness of bedside fetal fibronectin testing varies according to treatment algorithm. <i>J Matern Fetal Med</i> 2001; <b>10</b> :380–4	Test
Sunagawa S, Takagi K, Ono K, Miyachi K, Kikuchi A. Comparison of biochemical markers and cervical length for predicting preterm delivery. <i>J Obstet Gynaecol Res</i> 2008; <b>34</b> :812–19. <a href="https://doi.org/10.1111/j.1447-0756.2008.00844.x">https://doi.org/10.1111/j.1447-0756.2008.00844.x</a>	Test
Swamy GK, Simhan HN, Gammill HS, Heine RP. Clinical utility of fetal fibronectin for predicting preterm birth. <i>J Reprod Med</i> 2005; <b>50</b> :851–6	Test
Tanir HM, Sener T, Yildiz Z. Cervicovaginal fetal fibronectin (FFN) for prediction of preterm delivery in symptomatic cases: a prospective study. <i>Clin Exp Obstet Gynecol</i> 2008; <b>35</b> :61–4	Test

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
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. <i>Ultrasound Obstet Gynecol</i> 2005; <b>26</b> :699–706. <a href="https://doi.org/10.1002/uog.2633">https://doi.org/10.1002/uog.2633</a>	Test
Tekesin I, Marek S, Hellmeyer L, Reitz D, Schmidt S. Assessment of rapid fetal fibronectin in predicting preterm delivery. <i>Obstet Gynecol</i> 2005; <b>105</b> :280–4	Test
Tekesin I, Wallwiener D, Schmidt S. The value of quantitative ultrasound tissue characterization of the cervix and rapid fetal fibronectin in predicting preterm delivery. <i>J Perinat Med</i> 2005; <b>33</b> :383–91. <a href="https://doi.org/10.1515/JPM.2005.070">https://doi.org/10.1515/JPM.2005.070</a>	Test
Triana H, Saldivar D, Contreras N, Guzman A, Soria A, Farran I, et al. Fetal Fibronectin in the Diagnosis of Preterm Labor. In Cabero L, Carrera JM, editors. <i>Perinatology 2001 Volumes 1 and 2</i> . Bologna: Monduzzi International Proceedings Division; 2001. pp. 599–602	Test
Tsoi E, Akmal S, Geerts L, Jeffery B, Nicolaides KH. Sonographic measurement of cervical length and fetal fibronectin testing in threatened preterm labor. <i>Ultrasound Obstet Gynecol</i> 2006; <b>27</b> :368–72. <a href="https://doi.org/10.1002/uog.2723">https://doi.org/10.1002/uog.2723</a>	Test
Tsoi E, Akmal S, Geerts L, Jeffery B, Nicolaides KH. Sonographic measurement of cervical length and fetal fibronectin testing in threatened preterm labor. <i>Obstet Gynecol Surv</i> 2006; <b>61</b> :441–2	Test
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). <i>Acta Obstetrica et Gynecologica Scandinavica</i> 2009; <b>88</b> :1319–30. URL: <a href="http://onlinelibrary.wiley.com/doi/10.1111/j.1365-3113.2009.04519.x">http://onlinelibrary.wiley.com/doi/10.1111/j.1365-3113.2009.04519.x</a> (accessed September 2017)	Test
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. <i>Ultrasound Obstet Gynecol</i> 2018; <b>51</b> :596–603	Test
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. <i>Am J Obstet Gynecol</i> 2014; <b>210</b> (Suppl. 1):S375	Test
Voluménie JL, Guibourdenche J, Doridot V, Sibony O, Oury JF, Blot P, Luton D. Failure of cervical fibronectin to predict premature delivery in a population of monofetal pregnancies with idiopathic preterm labor. <i>Eur J Obstet Gynecol Reprod Biol</i> 2001; <b>97</b> :35–9	Test
von Schöning D, Fischer T, von Tucher E, Slowinski T, Weichert A, Henrich W, Thomas A. Cervical sonoelastography for improving prediction of preterm birth compared with cervical length measurement and fetal fibronectin test. <i>J Perinat Med</i> 2015; <b>43</b> :531–6. <a href="https://doi.org/10.1515/jpm-2014-0356">https://doi.org/10.1515/jpm-2014-0356</a>	Test
Watson DL, Kim SJ, Humphrey MD. Study of cervicovaginal fetal fibronectin status to guide treatment of threatened preterm labour. <i>Aust N Z J Obstet Gynaecol</i> 1998; <b>38</b> :185–7	Test
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. <i>Ultrasound Obstet Gynecol</i> 2017; <b>24</b>	Test
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. <i>Eur J Obstet Gynecol Reprod Biol</i> 2015; <b>192</b> :47–53 <a href="https://doi.org/10.1016/j.ejogrb.2015.06.008">https://doi.org/10.1016/j.ejogrb.2015.06.008</a>	Test
Wilms FF, Vis JY, de Wit-Zuurendonk L, Porath MM, Mol BW. What is the risk of preterm delivery after arrested preterm labor? <i>Am J Perinatol</i> 2015; <b>32</b> :63–70. <a href="https://doi.org/10.1055/s-0034-1374818">https://doi.org/10.1055/s-0034-1374818</a>	Test
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. <i>Clin Chem</i> 2007; <b>53</b> :1534–40	Test
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. <i>J Obstet Gynaecol Res</i> 2007; <b>33</b> :38–44	Test

continued

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Yoneda S, Shiozaki A, Yoneda N, Shima T, Ito M, Yamanaka M, <i>et al.</i> 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. <i>J Obstet Gynaecol Res</i> 2011; <b>37</b> :861–6. <a href="https://doi.org/10.1111/j.1447-0756.2010.01453.x">https://doi.org/10.1111/j.1447-0756.2010.01453.x</a>	Test
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? <i>Aust N Z J Obstet Gynaecol</i> 2009; <b>49</b> :16–21	Test
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 $\leq$ 20 mm. <i>J Matern Fetal Neonatal Med</i> 2017; <b>30</b> :294–7. <a href="https://doi.org/10.3109/14767058.2016.1171309">https://doi.org/10.3109/14767058.2016.1171309</a>	Test
Management of preterm labour. Volume 1: Evidence report and appendices; Volume 2: Evidence tables (Structured abstract). <i>Health Technol Assess</i> 2000; <b>1</b> (4). URL: <a href="http://onlinelibrary.wiley.com/doi/10.1002/hta.32001000946/frame.html">http://onlinelibrary.wiley.com/doi/10.1002/hta.32001000946/frame.html</a> (accessed September 2017)	Unobtainable
Fetal fibronectin for the prediction of preterm labor (Structured abstract). <i>Health Technol Assess</i> 2000(4).	Unobtainable
Fetal fibronectin point of care test (project record). <i>Health Technol Assess</i> 2006(2). URL: <a href="http://onlinelibrary.wiley.com/doi/10.1002/hta.32006000075/frame.html">http://onlinelibrary.wiley.com/doi/10.1002/hta.32006000075/frame.html</a>	Unobtainable
Actim Partus test (project record). <i>Health Technol Assess</i> 2010(4). URL: <a href="http://onlinelibrary.wiley.com/doi/10.1002/hta.32010001243/frame.html">http://onlinelibrary.wiley.com/doi/10.1002/hta.32010001243/frame.html</a>	Unobtainable
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. <i>Int J Gynecol Obstet</i> 2009; <b>107</b> :S413	Unobtainable
American College of Obstetricians and Gynecologists. Fetal fibronectin preterm labor risk test. <i>Int J Gynecol Obstet</i> 1997; <b>59</b> :164	Unobtainable
Anonymous. Fetal fibronectin measurement to predict preterm delivery. <i>Am Fam Physician</i> 1992; <b>45</b> :287	Unobtainable
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. <i>IJEES</i> 2015; <b>5</b> :29–34	Unobtainable
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. <i>Int J Gynecol Obstet</i> 2012; <b>119</b> :S291	Unobtainable
Bertani D, Beski L, Tagliavini M, Benassi L. <i>Prognostic Role of Fetal Fibronectin in Threatened Preterm Delivery</i> . Merialdi A, Cavatorta E, Gramellini D, editors. Parma: 18 <sup>th</sup> Meeting on Prenatal Medicine; 1993	Unobtainable
Bianco M, Depew C, Meisel R. A fetal fibronectin assay for detection of rupture of membranes. <i>Am J Obstet Gynecol</i> 1993; <b>168</b> :435	Unobtainable
Burwick R, York N, Lee, Ross M, Kjos S. Cervical assessment of cervical length is equivalent to fetal fibronectin in the prediction of preterm delivery. American College of Obstetricians and Gynecologists' Annual Meeting, San Francisco, CA, 17–19 May 2010. URL: <a href="http://onlinelibrary.wiley.com/doi/10.1002/hta.3200755741/frame.html">http://onlinelibrary.wiley.com/doi/10.1002/hta.3200755741/frame.html</a> (accessed September 2017)	Unobtainable
Crane J, Armson BA, Dodds L, Kennedy W, Feinberg R. Predictors of preterm delivery. <i>Am J Obstet Gynecol</i> 1998; <b>178</b> :S189	Unobtainable
Gervasi MT, Marsoni V, Reveane A, Serra MM, Bracalente G, Darin G. <i>Role of Fetal Fibronectin in the Diagnosis of Preterm Delivery</i> . Cosmi EV, editor. Spoleto: 6 <sup>th</sup> National Congress–Societa Italiana Di Medicina Perinatale; 1996	Unobtainable
Ghidini A, Poggi SH, Korker V. Performance of vaginal fetal fibronectin as predictor of preterm delivery in a community hospital. <i>J Soc Gynecol Investig</i> 2006; <b>13</b> :338A-A	Unobtainable
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. <i>Geburtshilfe Und Frauenheilkunde</i> 2008; <b>68</b> :S25-S	Unobtainable
Hayes, Inc. Fetal fibronectin test in women with symptoms of preterm labor (Structured abstract). <i>Health Technol Assess</i> 2006(4). URL: <a href="http://onlinelibrary.wiley.com/doi/10.1002/hta.32006001571/frame.html">http://onlinelibrary.wiley.com/doi/10.1002/hta.32006001571/frame.html</a> (accessed September 2017)	Unobtainable

TABLE 31 Studies excluded at full-text review, with reasons (continued)

Citation	Reason for exclusion
Klee GG, Arnold SE, Heise RH. Fetal fibronectin as a risk predictor of preterm delivery. <i>Clinical Chemistry</i> 2000; <b>46</b> :A65-A	Unobtainable
Lu GC, Goldenberg RL, Cliver SP, Kreaden US, Andrews WW. Quantitative fetal fibronectin levels and the prediction of spontaneous preterm birth in symptomatic women. <i>Am J Obstet Gynecol</i> 2000; <b>182</b> :S53	Unobtainable
Luzzi VI, Hankins K, Gronowski AM. Predictive value of fetal fibronectin as a biochemical marker of preterm delivery using the rapid fFN TLiTM system. <i>Clinical Chemistry</i> 2002; <b>48</b> :A121-A	Unobtainable
Macones GA, Stamilio D. Fetal fibronectin testing: the cost for false negative/positive results across gestational ages. <i>Am J Obstet Gynecol</i> 2000; <b>182</b> :S46	Unobtainable
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. <i>Clinical Chemistry</i> 1999; <b>45</b> :A94-A	Unobtainable
Mims O, Taylor S, Sundaram S, Stevens M, Manimekalai S, Goldstein PJ. Fetal fibronectin assay to predict pre term labor. <i>Clinical Chemistry</i> 1999; <b>45</b> :A72-A	Unobtainable
Min DL, Haberman S, McCalla S. The impact of the rapid fetal fibronectin assay on the management of preterm labor. <i>Obstetrics and Gynecology</i> 2004; <b>103</b> :73S-5	Unobtainable
Morgan BR. Clinical performance of the Adeza biomedical Tli Rapid Fetal Fibronectin Test at a multispecialty clinic. <i>Clinical Chemistry</i> 2001; <b>47</b> :A145-A	Unobtainable
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). <i>Int J Gynecol Obstet</i> 2009; <b>107</b> :S468	Unobtainable
Oyelese Y, Peters S, Ananth CV, Kinzler W, Smulian JC. The utility of fetal fibronectin in predicting preterm birth in clinical practice. <i>J Soc Gynecol Investig</i> 2006; <b>13</b> :123A-A	Unobtainable
Paternoster DM, Lazzarin L, Stella A, Pignataro R. <i>Fetal Fibronectin and Determination of Blood Ferritin Levels in Preterm Delivery</i> . Cosmi EV, editor. Spoleto: 6 <sup>th</sup> National Congress–Societa Italiana Di Medicina Perinatale; 1996	Unobtainable
Pates JA, Lattu AL, Nielsen PE. The value of fetal fibronectin in predicting the onset of labor before 41 weeks of gestation. <i>Obstetrics and Gynecology</i> 2008; <b>111</b> :102S-5	Unobtainable
Paulson MC, Lapointe J, Sheldon E, Mason R, Hernandez O. Elimination of sample warming for rapid fFN testing. <i>Clinical Chemistry</i> 2002; <b>48</b> :A194-A	Unobtainable
Paulson MC, Sheldon EL. A Comparison of rapid fFN for the Tli system and the rapid fFN for the Tli IQ system. <i>Clinical Chemistry</i> 2001; <b>47</b> :A111-A	Unobtainable
Quintieri F, Paternoster D, Depaoli M, Grella PV, Plebani M. <i>Fetal Fibronectin in Early Diagnosis of Preterm Births</i> . Merialdi A, Cavatorta E, Gramellini D, editors. Parma: 18 <sup>th</sup> Meeting on Prenatal Medicine; 1993	Unobtainable
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. <i>Am J Obstet Gynecol</i> 2000; <b>182</b> :S44	Unobtainable
Reveane A, Marsoni V, Gervasi MT, Amici G, Serra M, Marzolini M. <i>Correlation between fetal fibronectin levels and the risk of premature delivery</i> . Cosmi EV, editor. Spoleto: 6 <sup>th</sup> National Congress–Societa Italiana Di Medicina Perinatale; 1996	Unobtainable
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. <i>Am J Obstet Gynecol</i> 2000; <b>182</b> :S49	Unobtainable
Sheldon EL, Paulson MC, Hernandez O. The effect of yeast on rapid fFN for the Tli(tm) system. <i>Clinical Chemistry</i> 2001; <b>47</b> :A184-A	Unobtainable
Sheldon EL, Paulson MC, Lockhart T, Marzolf G. Analytical evaluation of Rapid Fetal Fibronectin (fFN) for the Adeza Tli((TM)) System. <i>Clinical Chemistry</i> 2000; <b>46</b> :A4-A	Unobtainable
Siddiq M, Haloob R. An audit of fetal fibronectin test. <i>Int J Gynecol Obstet</i> 2012; <b>119</b> :S752	Unobtainable
Volpogni C, Facchinetti F, Martinez F, Genazzani AR. <i>Fetal Fibronectin as Predictor of Preterm Birth</i> . In Merialdi A, Cavatorta E, Gramellini D, editors. Parma: 18 <sup>th</sup> Meeting on Prenatal Medicine; 1993	Unobtainable

## Raw rest accuracy data as reported in the papers

**TABLE 32** The DTA values for prediction of delivery within 48 hours

Study (first author and year)	N	True positive	False positive	True negative	False negative	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	+LR (95% CI)	-LR (95% CI)
<b>Actim Partus</b>											
Brik (2010) <sup>48</sup>	276	NR	NR	NR	NR	73.7	64.9	16.1	96.4	2.10 (1.52 to 2.91)	0.41 (0.19 to 0.87)
Goyal (2016) <sup>52</sup>	60	23	11	14	12	NR	NR	NR	NR	NR	NR
Lembet (2002) <sup>53</sup>	36	14	4	17	1	93.3	81	77.8	94.4	4.9 (2.0 to 11.9)	0.08 (0.01 to 0.50)
Ting (2007) <sup>56</sup>	94	NR	NR	NR	NR	100	74	18	100	NR	NR
Tripathi (2016) <sup>57</sup>	468	NR	NR	NR	NR	95.4	82.2	61.7	98.3	NR	NR
<sup>a</sup> Vishwekar (2017) <sup>58</sup>	30	11	3	14	2	73.3	64.3	68.8	69.2	NR	NR
<b>PartoSure</b>											
Werlen (2015) <sup>41</sup>	41	0	1	39	1	NR	NR	NR	NR	NR	NR

NR, not reported.

<sup>a</sup> One patient with a negative result absconded so delivery details could not be obtained.



**TABLE 33** The DTA values for prediction of delivery within 7 days

Study (first author and year) and test	N	True positive	False positive	True negative	False negative	Sensitivity, n/N (%) [95% CI]	Specificity, n/N (%) [95% CI]	PPV, n/N (%) [95% CI]	NPV, n/N (%) [95% CI]	+LR (95% CI)	-LR (95% CI)
Bruijn (2016) (APOSTEL-1) <sup>42,43</sup>											
ffN at < 10 ng/ml	350	66	162	119	3	NR	NR	NR	NR	NR	NR
ffN at < 200 ng/ml	49	46	46	235	20	NR	NR	NR	NR	NR	NR
ffN at < 500 ng/ml	29	12	12	269	40	NR	NR	NR	NR	NR	NR
Actim Partus	54	30	30	251	15	(78.3)	(89.3)	(64.3)	(94.4)	NR	NR
Hadzi-Lega (2017) <sup>44</sup>											
Actim Partus	57	5	12	39	1	5/6 (83) [35.88 to 99.58]	39/51 (76) [62.51 to 87.21]	5/17 (29) [10.31 to 55.96]	39/40 (98) [86.84 to 99.94]	NR	NR
PartoSure	5	5	5	46	1	5/6 (83) [35.88 to 99.58]	46/51 (90) [78.59 to 96.74]	5/10 (50) [18.71 to 81.29]	46/47 (98) [88.71 to 99.95]	NR	NR
Abo El-Ezz (2014) <sup>45</sup>											
Actim Partus	57	20	9	18	10	NR	NR	NR	NR	NR	NR
Altinkaya (2009) <sup>46</sup>											
Actim Partus	105	9	16	75	5	NR	NR	NR	NR	NR	NR
Azlin (2010) <sup>47</sup>											
Actim Partus	51	4	3	43	1	(80.0)	(93.5)	(57.1)	(97.7)	NR	NR
Brik (2010) <sup>48</sup>											
Actim Partus	276	NR	NR	NR	NR	73.1	66.2	21.8	95	2.16 [1.60 to 2.92]	0.41 [0.21 to 0.78]
Cooper (2012) <sup>49</sup>											
Actim Partus	349	2	89	254	4	2/6 (33) [0.00 to 0.71]	254/343 (74) [0.69 to 0.79]	2/91 (2) [0.00 to 0.05]	254/258 (98) [0.97 to 1.00]	1.28 [0.41 to 4.04]	0.90 [0.51 to 1.59]
Danti (2011) <sup>50</sup>											
Actim Partus	60	2	17	39	2	(50) [7 to 93]	(70) [56 to 81]	(11) [1 to 33]	(95) [83 to 99]	1.65 [0.57 to 4.74]	0.72 [0.27 to 1.94]
Eroglu (2007) <sup>51</sup>											
Actim Partus	51	5	7	38	1	(83.3)	(84.4)	(41.7)	(97.4)	5.36 [2.3 to 12.2]	0.20 [0.01 to 0.7]
Goyal (2016) <sup>52</sup>											
Actim Partus	60	26	8	8	18	NR	NR	NR	NR	NR	NR

continued

**TABLE 33** The DTA values for prediction of delivery within 7 days (*continued*)

Study (first author and year) and test	N	True positive	False positive	True negative	False negative	Sensitivity, n/N (%) [95% CI]	Specificity, n/N (%) [95% CI]	PPV, n/N (%) [95% CI]	NPV, n/N (%) [95% CI]	+LR (95% CI)	-LR (95% CI)
Lembet (2002) <sup>53</sup>											
Actim Partus	36	15	3	17	1	(93.8)	(85)	(83.3)	(94.1)	6.2 [2.2 to 17.8]	0.07 [0.01 to 0.5]
Riboni (2011) <sup>54</sup>											
Actim Partus	210	NR	NR	NR	NR	(50)	(83.7)	(10.8)	(97.7)	NR	NR
<sup>a</sup> Tanir (2009) <sup>55</sup>											
Actim Partus	68	14	11	42	1	14/15 (93.3)	42/53 (79.2)	14/25 (56)	42/43 (97.6)	4.4 [2.1 to 5.2]	0.8 [0.4 to 0.9]
Ting (2007) <sup>56</sup>											
Actim Partus	94	NR	NR	NR	NR	(69)	(78)	(39)	(92)	NR	NR
Tripathi (2016) <sup>57</sup>											
Actim Partus	468	NR	NR	NR	NR	(94.7)	(92.4)	(85.6)	(97.3)	NR	NR
<sup>b</sup> Vishwekar (2017) <sup>58</sup>											
Actim Partus	30	13	1	10	5	(68.4)	(90)	(92.9)	(60)	NR	NR
Bolotskikh (2017) <sup>59</sup>											
PartoSure	99	12	4	83	0	12/12 (100) [74 to 100]	83/87 (95) [89 to 99]	12/16 (75) [48 to 93]	83/83 (100) [96 to 100]	NR	NR
Nikolova (2015) <sup>61</sup>											
PartoSure	203	28	9	159	7	28/35 (80) [63.1 to 91.6]	159/168 (95) [90.1 to 97.5]	28/37 (76) [58.8 to 88.2]	159/166 (96) [91.5 to 98.3]	NR	NR
Werlen (2015) <sup>41</sup>											
PartoSure	41	0	1	39	1	(0) [0.0 to 9.75]	(97.5) [86.8 to 99.9]	(0) [0.0 to 97.5]	(97.5) [86.8 to 99.9]	NR	NR
Bruijn (2016) (Bruijn) <sup>62</sup>											
ffN at < 10 ng/ml	455	45	276	131	3	NR	NR	NR	NR	NR	NR
ffN at < 200 ng/ml	455	34	87	320	14	NR	NR	NR	NR	NR	NR
ffN at < 500 ng/ml	455	14	23	384	34	NR	NR	NR	NR	NR	NR

NR, not reported.

<sup>a</sup> 'When there were no visible lines, which was observed in two cases, a new sample was not taken. These patients were assigned test positive.'<sup>b</sup> One patient with a negative result absconded so delivery details could not be obtained.

## Appendix 4 Additional information on existing cost-effectiveness studies

### Further detail on observational cost-minimisation studies

The study by Abenhaim *et al.*<sup>136</sup> describes the cost implications for the addition of Rapid fFN to clinical examination in a tertiary university hospital in Montreal, Canada. The diagnostic protocol of 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 > 3 cm 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 was compared with clinical examination alone.

This study evaluated fFN in a setting in which 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 2 weeks (100% NPV); however, three of these women (9%) were admitted to hospital. The latter finding 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<sup>136</sup> acknowledged the absence of pharmaceutical and radiological and laboratory costs as a limitation of their study. In this regard, the study does not provide evidence on the proportion of women who were adequately managed with corticosteroid treatment (i.e. within 7 days of delivery) or, indeed, on the overall proportion of preterm deliveries including those beyond 2 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 who would have required additional evaluation if testing had been unavailable.

A US study<sup>137</sup> 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 from July 1995 to June 1997. 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 percentage of admissions from 28.1% to 17%, the mean number of 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. percentage of NICU admissions, median days of NICU length of stay, ventilation duration and percentage of 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 because it was based on a laboratory assay (as opposed to the Rapid fFN test commonly used these days) and required mothers testing negative to return 2 weeks later for fFN retesting.

### Other studies with relevant outcomes

Berghella and Saccone<sup>138</sup> systematically reviewed the RCT evidence on fFN and found that it resulted in an increasing trend towards admission to NICU (risk ratio 2.48, 95% CI 0.96 to 6.46) relative to clinical examination alone (blinded to fFN results), which had a 7.45% prevalence across the two RCTs reporting this outcome.<sup>76,139</sup>

## Detailed review of individual models

### Cost-minimisation studies

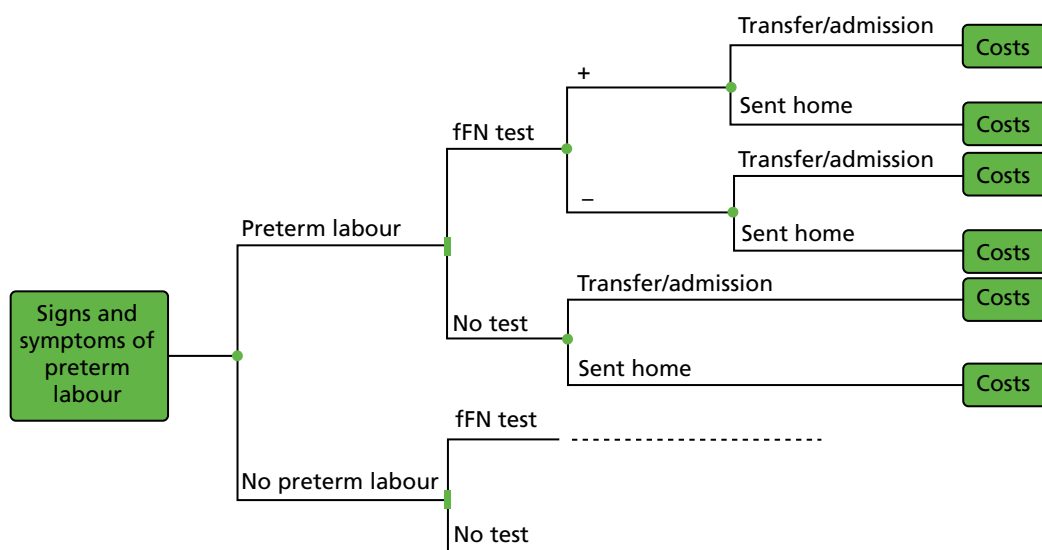
A decision analysis was used by Chuck and Nguyen<sup>72</sup> to evaluate the health system costs following the adoption of fFN testing in Alberta, Canada, in January 2008. Their evaluation used observational data from inpatient and outpatient administrative medical records covering the period from April 2002 to March 2013. It linked data from the provincial laboratory system to determine the proportion of patients 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 were 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 of patients who did not deliver before 37 weeks (*Figure 16*). For each of these subgroups, the decision between conducting the Rapid fFN test and not conducting 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 34*.

Chuck and Nguyen<sup>72</sup> 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 2008–13 period of observation. There were 69 more ambulance transfers and an additional 1379 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\$4M.

One limitation of Chuck and Nguyen's<sup>72</sup> 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.



**FIGURE 16** Decision tree of fFN testing strategy in Chuck and Nguyen.<sup>72</sup> Dotted lines denote the same tree structure as described for the preterm labour branch in the tree.

TABLE 34 Parameter values in Chuck and Nguyen<sup>72</sup>

Parameter	Parameter values				Source
	Inpatient database		Outpatient database		
	Estimate	95% CI	Estimate	95% CI	
<b>Not in preterm labour</b>					
<i>Transfers</i>					
Positive test results	0.32	0.26 to 0.39	0.14	0.11 to 0.18	Chuck and Nguyen <sup>72</sup>
Probability of transfer if not tested	0.06	0.05 to 0.07	0.03	0.02 to 0.04	
OR positive vs. not tested (OR+)	2.22	1.38 to 3.57	10.81	3.96 to 19.51	
OR negative vs. not tested (OR-)	0.78	0.51 to 1.19	1.53	0.81 to 2.88	
OR positive vs. negative	2.85	N/A	7.06	N/A	Calculations by AG = OR+/OR-
<i>Hospital admissions</i>					
Positive test results	UA	UA	0.11	0.09 to 0.12	
Probability of admission not tested	UA	UA	0.11	0.10 to 0.11	
OR positive vs. not tested (OR+)	UA	UA	5.38	3.65 to 7.95	
OR negative vs. not tested (OR-)	UA	UA	0.47	0.37 to 0.60	
OR positive vs. negative	UA	UA	11.44	N/A	
<b>In preterm labour</b>					
Positive test results	0.41	0.35 to 0.46	0.31	0.28 to 0.34	Chuck and Nguyen <sup>72</sup>
Probability of transfer if not tested	0.06	0.05 to 0.07	0.22	0.21 to 0.23	
OR positive vs. not tested (OR+)	7.45	3.89 to 14.27	3.68	2.55 to 5.31	
OR negative vs. not tested (OR-)	1.91	1.11 to 3.29	1.26	0.96 to 1.66	
OR positive vs. negative	3.90	N/A	2.92	N/A	Calculations by AG = OR+/OR-
Hospital admissions	UA	UA	Not reported	Not reported	Model assumes that no cost savings are realised
N/A, not applicable; UA, unavailable.					

The strengths of the study are found in its documenting of patient management consequent on test results, particularly in relation to transfers from lower level to tertiary units, and hospital admissions.

A UK study also modelled the cost difference between fFN plus clinical examination with clinical examination alone based on signs and symptoms.<sup>70</sup> Costs were measured for the time of hospital observation up to delivery, as the evaluated test strategies were assumed by the authors to not 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<sup>76</sup>) (Table 35). In terms of costs, the analysis used an activity-weighted average length of stay of *NHS Reference Costs* HRG NZ07 and NZ08 (for short and long stay), and the rate of hospital transfers and the proportion of tocolysis administered intravenously were assumed to be the same across arms, as was the number of ultrasound scans per admission ( $n = 1$ ).<sup>79</sup> Owing to a 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<sup>85</sup> and doses from the guidelines of the Royal College for Obstetricians and Gynaecologists.<sup>140,141</sup> The costs of hospital transfer and the cost of ultrasound scan (HRG 501OU) were obtained from *NHS Reference Costs*.<sup>79</sup> The study found that

**TABLE 35** Model parameter values in Deshpande *et al.*<sup>70</sup>

Parameter	Value	Standard error	Source
Admission rate with fFN (positive)	1.00	Not applicable (value fixed by assumption)	Reproduction from Dutta and Norman <sup>76</sup> by Deshpande <i>et al.</i> <sup>70</sup>
Incidence of tocolysis with fFN (positive)	0.286	$= (0.286 \times 0.714/7)^{0.5}$	
Incidence of steroids with fFN (positive)	0.714	$= (0.714 \times 0.286/7)^{0.5}$	
Transfer from hospital with fFN (positive)	0.167	$= (0.167 \times 0.833/6)^{0.5}$	
Admission rate with fFN (negative)	0.324	$= (0.324 \times 0.676/37)^{0.5}$	
Incidence of tocolysis with fFN (negative)	0.027	$= (0.027 \times 0.973/37)^{0.5}$	
Incidence of steroids with fFN (negative)	0.297	$= (0.297 \times 0.703/37)^{0.5}$	
Transfer from hospital with fFN (negative)	0.056	$= (0.056 \times 0.944/36)^{0.5}$	

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 health-care costs to the NHS.

### Cost-effectiveness studies

Boyd *et al.*<sup>73</sup> designed a decision tree model with the aim of informing the design of a non-inferiority RCT<sup>142</sup> of fFN that included measuring cost-effectiveness neonatal outcomes from the NHS perspective. The model measured the benefits of accurately diagnosing preterm 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'.<sup>73</sup> Unlike other models in this field, the authors of this model assigned a < 100% admission probability (93%) given positive test results, based on UK audit data.<sup>83,143</sup> They also adopted a 90% probability of admission in the clinical-examination-only arm, as a best guess assumption (*Table 36*). The model included the costs of

**TABLE 36** Key parameters from Boyd *et al.*<sup>73</sup>

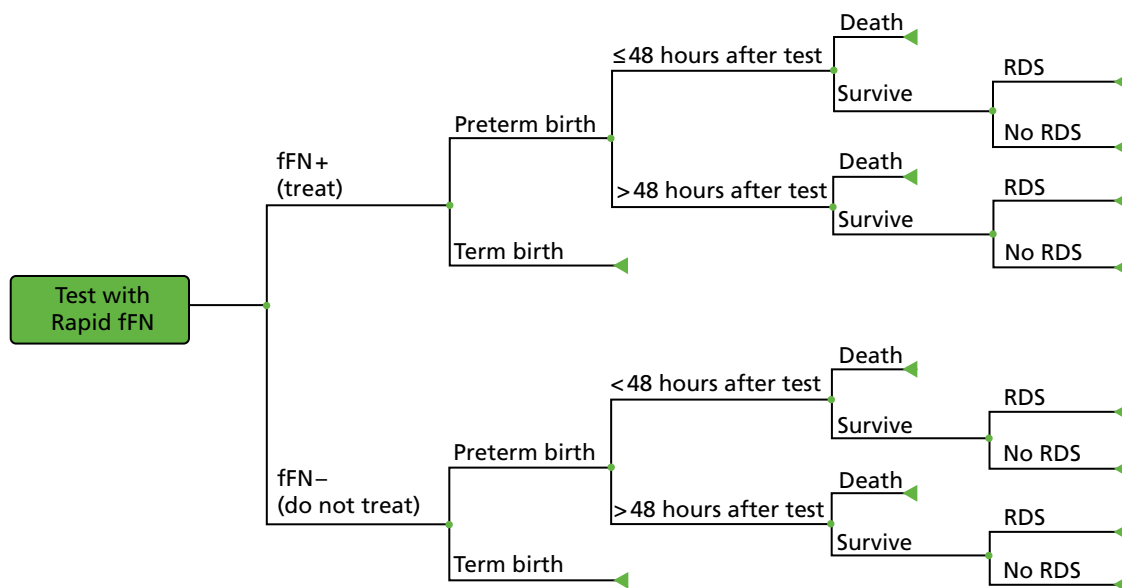
Model parameter	Value	Source and comments
Probability of preterm birth	0.20	Probability of PTB among TPL population in ORACLE II (Kenyon <i>et al.</i> <sup>144</sup> ); however, the figure is not found in the source
Probability of preterm morbidity	0.244	Admission to neonatal intensive care or specialised care (equivalent to BAPM level 1–3) in ORACLE II (Kenyon <i>et al.</i> <sup>144</sup> )
Steroid risk reduction	0.54	Relative risk preterm morbidity reduction (i.e. admission to intensive care or specialised care) with steroids (Roberts and Dalziel <sup>32</sup> )
Probability of death	0.0257	Probability of death in preterm births (ISD 2008 <sup>145</sup> )
Probability of hospital admission with fFN (positive)	0.93	Audit data (HC 2008 <sup>82</sup> and Hogg, Penney and Carmichael 2007 <sup>83</sup> )
Probability of hospital admission with fFN (negative)	0.33	Audit data (HC 2008 <sup>82</sup> and Hogg, Penney and Carmichael 2007 <sup>83</sup> )
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 <i>et al.</i> 2007 <sup>146</sup> )
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

CLU, consultant-led unit; PTB, preterm birth; TPL, threatened preterm labour.

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) owing to false-positive test results (Boyd *et al.*<sup>73</sup>). Furthermore, the model does not account for variation in costs and benefits by gestational age, thus ignoring the dramatically different implications of missing a premature birth, for example under 28 weeks versus older gestational ages. Furthermore, it did not measure negative effects of steroids use in false-positive cases and assumed that only preterm infants who received intensive or specialised care, 24% (in the ORACLE II RCT, Kenyon *et al.*<sup>144</sup>), 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'.<sup>73</sup>

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.<sup>74</sup> The study compared these strategies with 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 17*). 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 and (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 to not affect the probability of preterm delivery, and the model accounted for the reduced effects of tocolysis owing to 24-hour treatment as opposed to a 48-hour treatment course. The sensitivity and specificity of Rapid fFN in predicting preterm birth were used to populate the model. Relevant parameters from this analysis are presented in *Table 37*.

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



**FIGURE 17** Model of Rapid fFN testing from Mozurkewich *et al.*<sup>74</sup>

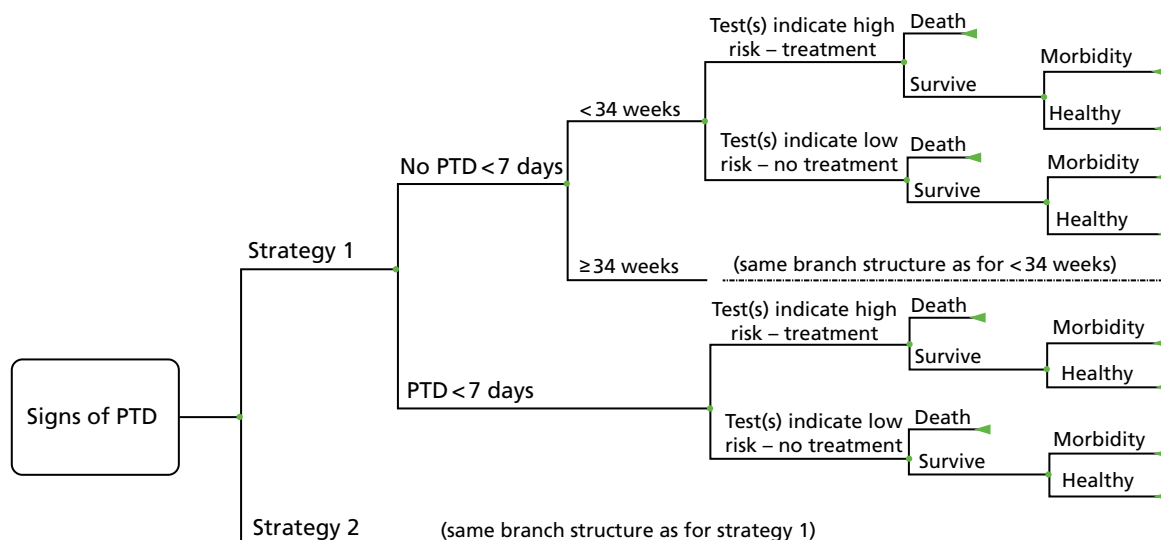
**TABLE 37** Effectiveness parameters based on Mozurkewich *et al.*<sup>74</sup>

Parameter	Value	Range	Source
Probability of preterm birth	0.50	0.20–0.71	Gyetvai <i>et al.</i> <sup>147</sup> and Moutquin <i>et al.</i> <sup>148</sup>
Baseline proportion of premature deliveries delivered within 48 hours	0.5	0.2–0.8	Moutquin <i>et al.</i> <sup>148</sup>
Effectiveness of tocolytics for delay of birth of < 48 hours	0.44	0.18–0.62	Gyetvai <i>et al.</i> <sup>147</sup>
Fractional decrease in effectiveness of tocolytics with short-term treatment	0.5	0.2–0.8	Authors' assumption (arbitrary owing to lack of data)
Probability of RDS	0.23	0.07–0.67	Crowley <i>et al.</i> <sup>149</sup>
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 <i>et al.</i> <sup>149</sup> on reduced effectiveness in terms of RDS)

transfers of women presenting with threatened preterm labour 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 owing to considerations other than the detection of fFN and, less frequently, some women testing positive may be discharged.<sup>73</sup>

A model for the Netherlands used information on treatment effects of tocolytic and steroid administration from the APOSTEL-II study<sup>150</sup> and diagnostic accuracy from APOSTEL-I,<sup>151</sup> reviewed in *Chapter 2*, to evaluate qualitative fFN, against a treat-all strategy.<sup>75</sup> The economic evaluation was undertaken 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 above grade II, periventricular leucomalacia above grade I and necrotising 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 (*Figure 18*) 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 a lack of data made projections highly uncertain. Key model parameters are presented in *Table 38*.





**FIGURE 18** Model of fFN diagnostic testing in van Baaren *et al.*<sup>75</sup> PTB, preterm delivery. Dotted line denotes the same tree structure as described for the strategy 1 branch in the tree.

**TABLE 38** Effectiveness parameters based on van Baaren *et al.*<sup>75</sup>

Parameter	Value	Range	Source
Probability of preterm delivery within 7 days of presentation	0.14	0.12–0.17	APOSTEL-I <sup>151</sup> and van Baaren <i>et al.</i> <sup>152</sup>
Preterm delivery > 7 days post testing and < 34 weeks' gestation	0.10	0.08–0.13	APOSTEL-I <sup>151</sup> and van Baaren <i>et al.</i> <sup>152</sup>
fFN positive in PTB within 7 days	0.90	0.82–0.94	APOSTEL-I <sup>151</sup> and van Baaren <i>et al.</i> <sup>152</sup>
fFN positive in PTB > 7 days	0.61	0.49–0.72	APOSTEL-I <sup>151</sup> and van Baaren <i>et al.</i> <sup>152</sup>
fFN positive in birth ≥ 34 weeks' gestation	0.37	0.33–0.41	APOSTEL-I <sup>151</sup> and van Baaren <i>et al.</i> <sup>152</sup>
Perinatal death with ACSs in PTB within 7 days	0.05	0.02–0.10	APOSTEL-II <sup>151</sup> and Roos <i>et al.</i> <sup>150</sup>
Perinatal death with ACSs in PTB after 7 days	0.04	0.02–0.08	APOSTEL-II <sup>151</sup> and Roos <i>et al.</i> <sup>150</sup>
Perinatal death in births ≥ 34 weeks' gestation	0.01	0.00–0.02	APOSTEL-II <sup>151</sup> and Roos <i>et al.</i> <sup>150</sup>
Severe adverse neonatal outcomes <sup>a</sup> with ACSs in PTB within 7 days	0.29	0.22–0.38	APOSTEL-II <sup>151</sup> and Roos <i>et al.</i> <sup>150</sup>
Severe adverse neonatal outcomes <sup>a</sup> with ACSs in PTB after 7 days	0.19	0.14–0.26	APOSTEL-II <sup>151</sup> and Roos <i>et al.</i> <sup>150</sup>
Severe adverse neonatal outcomes- in births ≥ 34 weeks' gestation	0.01	0.00–0.02	APOSTEL-II <sup>151</sup> and Roos <i>et al.</i> <sup>150</sup>
RR of perinatal death with corticosteroids	0.77	0.67–0.89	Roberts and Dalziel <sup>32</sup>
RR of severe adverse neonatal outcome <sup>a</sup> with corticosteroids within 7 days <sup>b</sup>	0.59	0.41–0.88	Roberts and Dalziel <sup>32</sup>

PTB, preterm birth; RR, risk reduction.

a Composite of IVH above grade II, chronic lung disease, PVL above grade I, necrotising enterocolitis, neonatal sepsis and perinatal mortality.

b The reporting of these parameters in van Baaren *et al.*<sup>152</sup> appears inconsistent, as no RR is presented for adverse neonatal events with corticosteroids in PTB (< 34 weeks) after 7 days.



## Appendix 5 Supplementary discussion and tables for the systematic review and selection of utilities

### Quality-of-life outcomes for preterm children

Studies concerning the quality-of-life outcomes of preterm children are summarised in *Table 39*. Of the 24 papers shortlisted, seven were deemed as lower priority, because 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.<sup>153</sup> A second study measures but does not report any SF-36 scores.<sup>103</sup> Four studies use quality-of-life measures that do not have mapping functions that allow for conversion to utilities.<sup>102,154–156</sup> Finally, another study measures utilities for 140 15- to 16-year-olds that had a gestational age pf < 29 weeks using the Health Utilities Index version 3 (HUI3);<sup>157</sup> however, this was only an abstract, and it did not report any values.

Of the remaining 17 papers (authors marked as bold in *Table 39*),<sup>103,110,114,161–167,170–175</sup> 12 provide direct measures of utility.<sup>110,161,164–167,170–175</sup> Seven of these studies use a version of the HUI.<sup>158–164</sup> One of these studies is a systematic review of quality of life and also reports utilities drawn from other sources.<sup>161</sup>

Three studies from Finland use the 17D measure.<sup>165–167</sup> Five studies reported means and SDs for the SF-36 measure of health-related quality of life.<sup>111–113,168,169</sup> The remaining two studies are modelling papers that make use of utilities drawn from other sources.<sup>109,170</sup>

Many studies in *Table 39* that were model based cited sources for the utilities they used. These source papers were collated and are summarised in *Table 40*. The majority of these studies used either the HUI2 domains, which had been converted into a utility using a multiattribute health status utility function,<sup>171</sup> or a direct utility measure based on standard gamble. The study by Carroll and Downs<sup>108</sup> and some of the studies reported in Tengs and Wallace<sup>172</sup> also used time trade-off methods of utility elicitation.

### Quality-of-life outcomes for mothers

Studies concerning the quality-of-life outcomes for mothers are summarised in *Table 41*. Of these four studies, one was not used because its measure of quality of life cannot be mapped into a utility value.<sup>105</sup> The remaining three studies report summary scores for the SF-36.<sup>104,106,173</sup>

**TABLE 39** Summary of papers from the systematic search, with information on quality of life for preterm children

Study (first author and year)	Population	Sample size	Country	QoL measure	Parameters provided	Comments
Bastek (2012) <sup>109</sup>	Preterm children with 34 weeks $\leq$ GA < 36 weeks	N/A (literature review)	USA (although 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. <sup>108,115</sup> Utilities for moderate persistent asthma and moderate cerebral palsy were used as proxies for RDS and adverse neurodevelopment, respectively
Båtsvik (2015) <sup>112</sup>	GA $\leq$ 28 weeks or BW $\leq$ 1000 g; assessed at mean age of 24 weeks; with/without severe disability	43 preterm + 43 control	Norway	SF-36	SF-36 dimension means	
Baumgardt (2012) <sup>153</sup>	Preterm with BW < 1250 g; surveyed at median age of 23 weeks	52 preterm + 75 control	Switzerland	SF-36	SF-36 dimension means plotted but not explicitly provided	Scores are also separated by sex
Beaudoin (2013) <sup>103</sup>	Preterm with BPD; with RDS; with no respiratory complications	426 with BPD + 852 RDS/preterm/term	Canada	SF-36 version 2	SF-36 results not reported	
Berbis (2012) <sup>154</sup>	Gestational age between 24 and 32 weeks; assessed at age 6–10 years	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 with French population norms
Bianco (2011) <sup>155</sup>	GA $\leq$ 29 weeks and/or BW $\leq$ 1500 g; treated/not treated with surfactant; assessed at 18 months	89 preterm with surfactant + 61 preterm, no surfactant + 145 term	Italy	TAPQOL	TAPQOL of children treated/not treated with surfactant	Abstract only

Study (first author and year)	Population	Sample size	Country	QoL measure	Parameters provided	Comments
Cooke (2004) <sup>111</sup>	Preterm VLBW infants, assessed at age 19–22 years (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 <i>et al.</i> (2007) <sup>168</sup>	Preterm and term children, assessed at age 31 years	126 preterm + 66 term	New Zealand	SF-36	SF-36 dimension means	
Einerson (2016) <sup>170</sup>	N/A – paper presents a cost-effectiveness model for cervical length screening	N/A – see source paper information in <i>Table 40</i>	USA (although 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. <sup>115,172,174</sup> It is not clear exactly which figures have been used, or how they may have been combined
Feingold (2002) <sup>102</sup>	BW < 1501 g and GA < 33 weeks, assessed at age 18–19 years	43 IVH 0–2, no cysts + 10 IVH 3–4 and/or cysts	USA	HRQoL from CDC	Means of four dimensions of HRQoL reported, separated into two IVH severity level groups (0–2, no PVL; and 3–4, with/without PVL)	Unclear how to derive utilities from the HRQoL means
Gray (2007) <sup>157</sup>	GA < 29 weeks, assessed at age 15–16 years	140 preterm + 108 control	UK	HUI3	No figures provided	Abstract only – relative differences are reported but not absolute utilities
Husby (2016) <sup>169</sup>	Preterm with BW ≤ 1500 g, assessed at age 23 years	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 (2011) <sup>156</sup>	32 weeks ≤ GA < 36 weeks, assessed at preschool age (2–5 years)	218 responders	The Netherlands	TAPQOL	TAPQOL dimension means are provided	Unclear how to derive utilities from the TAPQOL means. Study also reports prevalence of various behavioural problems

continued

**TABLE 39** Summary of papers from the systematic search, with information on quality of life for preterm children (*continued*)

Study (first author and year)	Population	Sample size	Country	QoL measure	Parameters provided	Comments
Korvenranta (2010) <sup>165</sup>	GA < 32 weeks or BW < 1501 g, assessed at age 4	1752	Finland	Utilities derived from 17D <sup>166</sup>	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 (2011) <sup>167</sup>	GA < 32 weeks or BW < 1501 g, assessed at age 5 years	568 preterm + 173 control	Finland	Utilities derived from 17D <sup>166</sup>	Implied utilities (QALYs/5) for preterm/VLBW infants compared with controls	Being born in a level III hospital increased median QALYs by 0.03/5 = 0.006, relative to a level II hospital
Lund (2012) <sup>113</sup>	Preterm with BW ≤ 1500 g, and another group SGA. Assessment at age 20 years	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 (2009) <sup>158</sup>	20 weeks ≤ GA < 25 weeks, assessed at age 11 years (EPICure study)	190 preterm + 141 control	UK and Ireland	HUI3	HUI3 multiattribute 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 multiattribute utilities using methods from two studies <sup>175-176</sup>
Rautava (2009) <sup>166</sup>	BW ≤ 1500 g or GA < 32 weeks, assessed at age 5	588 preterm + 176 control	Finland	17D	Figures identical to those used in Lehtonen <i>et al.</i> <sup>167</sup>  Additional utility provided for live-born VLBW infants	Being born in a level III hospital (relative to a level II hospital) increased the mean QALYs by 0.5/5 = 0.1
Roberts (2013) <sup>159</sup>	GA < 28 weeks or BW < 1000 g, 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 or not the HUI3 score reported here is computed in the same way as the multiattribute score in Petrou <i>et al.</i> <sup>158</sup>
van Lunenburg (2013) <sup>160</sup>	GA < 32 weeks or BW < 1500 g, assessed at ages 19 and 28 (POPS cohort)	314	The Netherlands	HUI3	HUI3 multiattribute utility given at age 19 years, and at 28 years	Multiple imputation values are based on an algorithm (multivariate imputation by chained equations) that incorporates information from respondents to interpolate missing data values

Study (first author and year)	Population	Sample size	Country	QoL measure	Parameters provided	Comments
Verrips (2008) <sup>161</sup>	BW ≤ 1000 g included from three separate cohorts, assessed at age 12–16 years	150 (Canada), 65 (Germany), 126 (the Netherlands)	Germany, Canada, the Netherlands	HUI3	HUI3 multiattribute utility for ELBW in Canada, Germany and the Netherlands	Utility function based on Furlong <i>et al.</i> <sup>176</sup>
Wolke (2016) <sup>162</sup>	ELBW (Canada), VP or VLBW (Germany, the Netherlands), assessed at adolescence (12–16 years), early adulthood (19–26 years) and adulthood (> 26 years)	169 (Canada), 91 (Germany), 140 (the Netherlands)	Germany, Canada, the Netherlands	HUI3	HUI3 utilities reported for three life stages for Canada, Germany and the 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 was measured at both early adulthood and adulthood
Wolke (2013) <sup>163</sup>	BW < 1500 g or GA < 32 weeks, assessed at age 13 years	260 preterm + 282 control	Germany	HUI3	HUI3 multiattribute utilities for VP/VLBW infants and controls, reported by both parents and children	Paper also reports other social and cognitive characteristics. In particular, mean IQ in VP/VLBW infants = 92, versus 101 in full-term controls  Another group of VP/VLBW infants who could not report their own utility had a parent-reported value of 0.18, but only based on a small sample ( <i>n</i> = 12)
Zwicker (2008) <sup>164</sup>	VLBW or preterm	N/A – see <i>Table 40</i> 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 QoL scores from other studies. Utility scores reported are from multiple studies <sup>177–181</sup>

17-dimensional Health-related Quality of Life; BPD, bronchopulmonary dysplasia; BW, birthweight; ELBW, extremely low birthweight; GA, gestational age; HRQoL, health-related quality of life; IQ, intelligence quotient; POPS, Project On Preterm and Small for gestational age infants; QoL, quality of life; SGA, small for gestational age; TAPQOL, TNO-AZL Preschool children Quality of Life; VLBW, very low birthweight; VP, very premature; VSP-A, Vécu et Santé Perçue des Adolescents.

**Note**

Green shading represents studies that did not provide any appropriate quality-of-life utility measures.

**TABLE 40** Secondary sources for utilities (referenced by papers found in the systematic search)

Study (first author and year)	Sample	Country	Utility measure	Utilities reported	Comments
Carroll (2009) <sup>108</sup>	4016 (> 18 years, 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 (2003) <sup>174</sup>	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 (1994) <sup>177</sup>	156 ELBW + 145 controls	Canada	HUI2 transformed using a MAHS utility function	Utilities are provided for different attribute score combinations on a six-dimension quality-of-life scale, for children at age 8 years	Originally excluded in full-text screening as it focused on low birthweight only  Utilities provided are for a combination of subjective health states, and do not correspond to any particular condition
Saigal (1994) <sup>178</sup>	156 ELBW + 145 controls	Canada	HUI2 transformed using a 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 (1996) <sup>179</sup>	141 ELBW + 145 controls	Canada	HUI2 (actual and for hypothetical states), standard gamble	Mean utilities for age 12–16 years reported, for own health states as well as four hypothetical scenarios	Sensitivity analysis results also provided. Provides average utility for ELBW, but not for specific medical conditions in isolation
Saigal (2000) <sup>180</sup>	149 (parents of ELBW) + 126 (parents of controls)	Canada	HUI2 (actual states), visual analogue scale and standard gamble (hypothetical states)	Parental perspectives of child's (12–16 years) utility, and perspectives for four 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 (2006) <sup>181</sup>	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



Study (first author and year)	Sample	Country	Utility measure	Utilities reported	Comments
Tengs (2000) <sup>172</sup>	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 (1999) <sup>115</sup>	42 (12 obstetricians, 15 pregnant women, 15 mothers)	The Netherlands	Standard gamble	Four health states: healthy child; transient neurological symptoms; permanent neurological symptoms; neonatal death. Each outcome has three 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

ELBW, extremely low birthweight; MAHS, multiattribute health status.

**TABLE 41** Summary of papers on quality of life for mothers

Study (first author and year)	Population	Sample size	Country	QoL measure	Parameters provided	Comments
Alemdaroglu (2009) <sup>104</sup>	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 (2009) <sup>106</sup>	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) <sup>173</sup>	Random sample of mothers of students in each of four different age groups (< 5, 5–10, 11–13, 14–18 years)	234	USA	SF-36 v2	SF-36 dimension means	Also provides mean SF-36 measures for mothers after splitting them into three age groups
Hill (2007) <sup>105</sup>	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 subdimensions also reported. Unclear how to convert this measure to a utility

LBW, low birthweight; QoL, quality of life.

**Note**

Green shading represents studies that did not provide suitable quality-of-life information for the PenTAG model.

## Summary tables containing raw and mapped utilities from all relevant studies identified in the systematic review of utilities

**TABLE 42** Summary of utilities for preterm children

Study (first author and year)	Population	Measure	Variable	Utility
Bastek (2012) <sup>109</sup>	34 weeks $\leq$ GA < 36 weeks	Standard gamble, TTO	Acute respiratory disease	0.87
			Chronic respiratory disease	0.88
			Neurodevelopmental delay	0.76
			Death	0.01
Båtsvik (2015) <sup>112</sup>	GA $\leq$ 28 weeks or BW $\leq$ 1000 g; 43 preterm + 43 control	SF-36 (assessed at mean age of 24 years)	Severe disability	0.763
			Healthy	0.846
Cooke (2004) <sup>111</sup>	Preterm (VLBW) infants; 79 preterm + 71 control	SF-36 (assessed at age 19–22 years)	All preterm	0.879
			Male	0.907
			Female	0.856
Dalziel (2007) <sup>168</sup>	Preterm and term children; 126 preterm + 66 control	SF-36 (assessed at age 31 years)	All preterm	0.887
Einerson (2016) <sup>170</sup>	Preterm survivors	Standard gamble, TTO	Neonatal death	0
			Severe neonatal morbidity	0.55
			Healthy neonate	1
Husby (2016) <sup>169</sup>	Preterm with BW $\leq$ 1500 g; 35 preterm + 37 control	SF-36 (assessed at age 23 years)	All preterm	0.857
			VLBW with no cerebral palsy, and not with low IQ	0.891

continued

**TABLE 42** Summary of utilities for preterm children (*continued*)

Study (first author and year)	Population	Measure	Variable	Utility
Korvenranta (2010) <sup>165</sup>	GA < 32 weeks or BW < 1501 g; n = 1752	17D (assessed at age 4 years)	None of the studied morbidities	0.9475
			23 weeks' GA	0.9025
			24–25 weeks' GA	0.9075
			26–27 weeks' GA	0.9175
			28–29 weeks' GA	0.9275
			30–31 weeks' GA	0.94
			≥ 32 weeks' GA	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
Two or more of the above morbidities	0.87			
Lehtonen (2011) <sup>167</sup> and Rautava (2009) <sup>166</sup>	GA < 32 weeks or BW < 1501 g; 568 preterm + 173 control	17D (assessed at age 5 years)	All preterm	0.92 (median)
				0.72 (mean)
			Live-born	0.94 (median)
				0.82 (mean)

Study (first author and year)	Population	Measure	Variable	Utility
Lund (2012) <sup>113</sup>	Preterm with BW ≤ 1500 g, and SGA; 43 VLBW + 55 SGA + 73 control	SF-36 (assessed at age 20 years)	All preterm	0.901
			SGA	0.888
Petrou (2009) <sup>158</sup>	20 weeks ≤ GA < 25 weeks; 190 preterm + 141 control	HUI3 (assessed at age 11 years)	All preterm	0.789
			≤ 23 weeks	0.772
			24 weeks	0.717
			25 weeks	0.83
Roberts (2013) <sup>159</sup>	GA < 28 weeks or BW < 1000 g; 194 preterm + 148 control	HUI3 (assessed at age 18 years)	All preterm	0.93
van Lunenburg (2013) <sup>160</sup>	GA < 32 weeks or BW < 1500 g; <i>n</i> = 314	HUI3	Age 19 years (assessed)	0.89
			Age 19 years (multiple imputation)	0.83
			Age 28 years (assessed)	0.88
			Age 28 years (multiple imputation)	0.85
Verrips (2008) <sup>161</sup>	BW ≤ 1000 g; 150 (Canada), 65 (Germany), 126 (the Netherlands)	HUI3 (assessed at 12–16 years)	Canada	0.76
			Germany	0.752
			The Netherlands	0.868
Wolke (2016) <sup>162</sup>	ELBW (Canada, <i>n</i> = 169); VP or VLBW (Germany <i>n</i> = 91, the Netherlands <i>n</i> = 140)	HUI3, assessed at adolescence (12–16 years), early adulthood (19–26 years), adulthood (> 26 years)	Canada, all preterm	0.79 (12–16 years)
				0.79 (19–26 years)
				0.73 (> 26 years)
			ELBW without neurosensory impairment (Canada)	0.83 (12–16 years)
				0.83 (19–26 years)
				0.77 (> 26 years)
			ELBW with neurosensory impairment (Canada)	0.68 (12–16 years)
	0.65 (19–26 years)			
	0.60 (> 26 years)			

continued

**TABLE 42** Summary of utilities for preterm children (*continued*)

Study (first author and year)	Population	Measure	Variable	Utility
			Germany, all preterm	0.82 (12–16 years)
				0.82 (19–26 years)
			The Netherlands, assessed	0.87 (12–16 years)
				0.89 (19–26 years)
				0.88 (> 26 years)
			The Netherlands, imputed	0.83 (19–26 years)
				0.85 (> 26 years)
Wolke (2013) <sup>163</sup>	GA < 32 weeks or BW < 1500 g; 260 preterm + 282 control	HUI3 (assessed at age 13 years)	Parent reported	0.88
			Child self-reported	0.84
Zwicker (2008) <sup>164</sup>	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

ELBW, extremely low birthweight; GA, gestational age; SGA, small for gestational age; TTO, time trade-off; VLBW, very low birthweight; VP, very premature.

**TABLE 43** Summary of utilities from secondary sources cited by studies found in the systematic search

Study (first author and year)	Population	Measure	Variable	Utilities
Carroll (2009) <sup>108</sup>	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 (2003) <sup>174</sup>	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 (1994) <sup>177</sup>	156 ELBW + 145 controls	HUI2 transformed using a MAHS utility function	All ELBW infants	0.82 (mean); 0.88 (median)
Saigal (1996) <sup>179</sup>	141 ELBW + 145 controls	Standard gamble (chance board)	All ELBW infants	0.87 (mean); 1.00 (median)
Saigal (2000) <sup>180</sup>	149 (parents of ELBW) + 126 (parents of controls)	Standard gamble (chance board)	All ELBW infants (parental assessments)	0.91 (mean); 1.0 (median)

continued

**TABLE 43** Summary of utilities from secondary sources cited by studies found in the systematic search (*continued*)

Study (first author and year)	Population	Measure	Variable	Utilities
Saigal <i>et al.</i> (2006) <sup>181</sup>	149 (parents of ELBW) + 126 (parents of controls)	Standard gamble	All ELBW infants	0.85 (mean, ELBW); 0.95 (median, ELBW)
			ELBW infants with neurosensory impairments	0.85
			ELBW infants without neurosensory impairments	0.85
Tengs and Wallace (2000) <sup>172</sup>	140 experts	Standard gamble	Cerebrovascular disease, intracranial aneurysm, good but incomplete recovery, normal life with minor neurological 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 owing to mental and/or physical disability	0.26
	156 patient proxies	HUI	ELBW infants (501–1000 g), assessed at age 8 years	0.82
	24 patients	Standard gamble	Lung disease, chronic (e.g. chronic bronchitis, emphysema, cystic fibrosis, on waiting list for transplant)	0.65
Lung disease, chronic (e.g. chronic bronchitis, emphysema, cystic fibrosis, transplant)			0.8	



Study (first author and year)	Population	Measure	Variable	Utilities
Vandenbussche <i>et al.</i> (1999) <sup>115</sup>	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)

ELBW, extremely low birth weight; ICU, intensive care unit; MAHS, multiattribute health status; TTO, time trade-off.

**TABLE 44** Mapped utilities for the quality of life of mothers

Study (first author and year)	Population	<i>n</i>	Measure	Variable	Utilities
Alemdaroglu (2015) <sup>104</sup>	Mothers of LBW infants, premature children with/without ICH or IVH	12	SF-36 (physical and mental summary only)	Mothers of children with ICH/IVH	1.021 <sup>a</sup>
		12		Mothers of children without ICH/IVH	1.016 <sup>a</sup>
Couto (2009) <sup>106</sup>	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) <sup>173</sup>	Random sample of mothers of students in each of four different age groups (< 5, 5–10, 11–13, 14–18 years)	234	SF-36 version 2	All mothers	0.640
		69		Mothers aged 25–34 years	0.529 <sup>b</sup>
		110		Mothers aged 35–44 years	0.525 <sup>b</sup>
		40		Mothers aged 45–54 years	0.532 <sup>b</sup>

LBW, low birthweight.

a These mapped utilities are greater than 1, suggesting that the mapping function is extrapolating too far outside the domain of the original sample used by Kim *et al.*<sup>182</sup>

b Utility mapped using a linear model (rather than quadratic), because SDs were not reported for the input SF-36 mean scores.

### Short Form questionnaire-36 items mapping and extraction of utilities

Although none of the studies that were found directly measured utilities based on the EQ-5D, various mapping functions exist that allow SF-36 summary measures to be converted into EQ-5D utilities. The EQ-5D is NICE's preferred measure of quality of life.<sup>114</sup> The University of Oxford Health Economics Research Centre maintains a database of such mapping studies.<sup>183</sup> The latest version was last updated in May 2016.

Two studies from this database were shortlisted, owing 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.<sup>101</sup> They show that this provides a more accurate prediction of EQ-5D utility than using ordinary least squares (OLS) estimates. Although censored models were also estimated, these are problematic to use owing to their non-linearity. Because only mean SF-36 data are 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 SDs 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. (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, while introducing bias. This is owing 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.) This version of the model was preferred to the linear model when 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 of 0–0.499.<sup>101</sup>

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.<sup>182</sup> 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 ( $R^2 = 0.6366$ , root square mean error = 0.16) was used to predict EQ-5D scores from the single study that reported outcomes for mothers of children with and without IVH or ICH, because this study did not report mean scores for each of the eight SF-36 dimensions.<sup>104</sup>

These SF-36-to-EQ-5D mapped utilities are tabulated in this appendix, along with the other relevant utilities extracted directly from papers (where available) as follows: *Table 42* contains all relevant utilities from studies of preterm children, *Table 43* contains all relevant utilities from studies that were identified as secondary sources of utility data and *Table 44* contains all relevant utilities from the studies of mothers.

## Study discussion

### Studies of 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.<sup>111,158</sup>

Two studies provide utilities for children born at different gestational ages.<sup>158,165</sup> Petrou *et al.*<sup>158</sup> only study children born between the gestational ages of 20 and 25 weeks, but assess utility using the HUI3 at 11 years of age in a UK and Ireland population. Korvenranta *et al.*<sup>165</sup> 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.

The largest studies are from Finland, and all three make use of the 17D quality-of-life measure.<sup>165–167</sup> These studies report QALYs rather than utilities, but because they are computed linearly, implied utilities are derived by dividing the QALY value by the overall time horizon (4 years in Korvenranta *et al.*,<sup>165</sup> 5 years in Lehtonen *et al.*<sup>167</sup>). No mapping studies from the 17D to the EQ-5D were found in the current University of Oxford Health Economics Research Centre database.<sup>183</sup> A 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.<sup>102</sup> The health-related quality-of-life measure they used, developed by the 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 years.

Likewise, only one paper considers quality of life for preterm children with RDS. The authors measure, but do not provide, SF-36 scores in their paper.<sup>103</sup> 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 *Utilities for respiratory distress syndrome*.

Although 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<sup>109</sup> identifies utilities (originally derived from two other studies<sup>108,115</sup>) for acute/chronic respiratory disease, and neurodevelopmental delay. The utilities from the study by Carroll and Downs<sup>108</sup> 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.<sup>108</sup> The utilities from this paper are used by Bastek *et al.*<sup>109</sup> for their model of ACSs, a treatment which is relevant to the economic model devised for this report. This study is discussed further in *Utilities for respiratory distress syndrome* and *Utilities for intraventricular haemorrhage*.

We received a forthcoming paper in confidence after contacting Dr Stavros Petrou (University of Warwick, 2017, personal communication), a health economist who has previously studied childhood outcomes.<sup>184</sup> 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 of > 18 years, which may be problematic when extrapolating over the entire lifespan.

### Studies of 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.<sup>104,105</sup> 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 low-birthweight preterm children with ICH or IVH and 12 who have low-birthweight 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,<sup>182</sup> it yields utilities of  $> 1$  (see *Table 44*) when applied to the data from Alemdaroglu *et al.*<sup>104</sup> 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.<sup>105</sup> 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<sup>173</sup> 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. Therefore, neither of these studies provide usable utilities for the economic model.

The final study of mothers, by Couto *et al.*,<sup>106</sup> captures the quality of life of mothers in Brazil who have had at least one of four previous adverse pregnancy outcomes. Although 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 who develop conditions but do not die and an overestimate for mothers of preterm children who 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.<sup>185</sup> We also consulted with Professor Dieter Wolke (University of Warwick, 2017, personal communication), 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.

### Utilities for respiratory distress syndrome

There is only one study identified that measured quality-of-life outcomes for preterm children with RDS.<sup>103</sup> The study compares four groups of subjects:

1. born preterm (gestational age  $< 37$  weeks) who developed bronchopulmonary dysplasia without infant RDS
2. born preterm with RDS but no subsequent bronchopulmonary dysplasia
3. born preterm without respiratory complications
4. born at term without respiratory complications.

The study uses a Canadian sample to whom questionnaires were administered containing the SF-36 version 2 via post. A total of 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 these 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 ACSs, Bastek *et al.*<sup>109</sup> utilise proxy utilities for acute and chronic respiratory disease. They argue that a utility value for a 10-day intensive care unit (ICU) admission was an acceptable proxy for

acute respiratory disease, because infants with RDS are managed in NICUs. The utility corresponding to this outcome is 0.87.<sup>108</sup> In *Table 43*, we see that Carroll and Downs<sup>108</sup> obtained this utility using the standard gamble method. Because NICE uses the UK time trade-off value set to obtain utilities from the EQ-5D, time trade-off utilities obtained by Carroll and Downs<sup>108</sup> are preferred to those obtained by standard gamble.<sup>114</sup> In the study by Carroll and Downs,<sup>108</sup> the time trade-off utility for 10-day ICU admission is 0.91.

Bastek *et al.*<sup>109</sup> report a utility for chronic respiratory disease of 0.88, which was taken from the study by Carroll and Downs<sup>108</sup> as the utility of moderate persistent asthma. In *Table 43*, we see that this utility was obtained using the standard gamble method. The time trade-off equivalent from the study by Carroll and Downs<sup>108</sup> is 0.91. 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 the study by Carroll and Downs.<sup>108</sup> This is 0.85, as elicited using the time trade-off method.

### Utilities for intraventricular haemorrhage

As discussed in *Study discussion*, only one study measured quality-of-life outcomes for children with IVH specifically.<sup>102</sup> This study uses four health-related quality-of-life questions from the CDC. One of these is measured on a 5-point scale, while 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.*<sup>109</sup> again provides the best available estimate of a utility value. They use the utility for moderate cerebral palsy from the study by Carroll and Downs<sup>108</sup> as a proxy for adverse neurodevelopment. In *Table 43*, we see that this utility of 0.76 is identical for both the standard gamble and time trade-off methods of elicitation.

### Utilities for preterm survivors

The NICE-preferred measure of health-related quality of life in adults is the EQ-5D;<sup>114</sup> however, given that we do not have any such data, the second-best option is to use SF-36 scores mapped onto the EQ-5D. Five studies provide utilities for children that were obtained from mapping the SF-36 mean dimension scores onto the EQ-5D.<sup>111–113,168,169</sup> Although 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.<sup>112,113,169</sup> One study restricts its population to gestation of  $\leq 28$  weeks or a birthweight of  $\leq 1000$  g.<sup>112</sup> This appears too limiting to capture the outcomes of all preterm children. The other two studies consider a more generous birthweight range of  $\leq 1500$  g, but have small preterm sample sizes of 35<sup>113</sup> and 43.<sup>169</sup>

Of the remaining two studies, one has SF-36 measures for a very low-birthweight sample of 79 participants from the UK (assessed between ages 19 and 22 years).<sup>111</sup> The other uses a larger sample of 126 preterm children from New Zealand (assessed at age 31 years) whose mothers were participants in the Auckland Steroid Trial.<sup>168</sup> Of these individuals, 21 had RDS in infancy, which may lead to a small downwards bias in the quality-of-life scores, although 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.

## Statistical analysis of the effects of birthweight on utility

In order to determine whether or not it was necessary to incorporate a utility reduction for lower birthweight, regression analysis was carried out on a data set obtained from one of the authors of a Canadian study.<sup>107</sup> The data contained utility, as measured by HUI3, assessed at three life stages (adolescence, young adulthood and mature adulthood). Along with this, we were supplied with data on birthweight, sex and gestational age. Data from 290 individuals were provided, although some had missing sex information and many participants did not respond in all three life stages. The mean birthweight in the sample was 2047.9 g (minimum = 560 g, maximum = 4734 g) and the mean gestational age was 33.2 weeks (minimum = 23 weeks, maximum = 40 weeks).

A random-effects GLS panel data estimator was used to estimate the following general model (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):

$$u_{it} = \alpha + \beta B_i + \gamma X_i + \theta D_t + \nu_i + \epsilon_{it}, \quad (14)$$

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,  $\nu$  = an unobservable fixed effect and  $\epsilon$  = the idiosyncratic error term. *Table 45* shows the estimates of five different model specifications.

**TABLE 45** Random-effects GLS estimates of utility on birthweight

Explanatory variable	Dependent variable: utility (HUI3)				
	(1)	(2)	(3)	(4)	(5)
Birthweight (g)	$9 \times 10^{-5}$ ( $4.96 \times 10^{-5}$ )	$9.07 \times 10^{-5*}$ ( $4.96 \times 10^{-5}$ )	0.00011 ( $9.06 \times 10^{-5}$ )	0.00011 ( $9.05 \times 10^{-5}$ )	0.0000292 ( $2.20 \times 10^{-5}$ )
Birthweight squared	$-1.08 \times 10^{-8}$ ( $1.10 \times 10^{-8}$ )	$-1.11 \times 10^{-8}$ ( $1.10 \times 10^{-8}$ )	$-1.4 \times 10^{-8}$ ( $1.52 \times 10^{-8}$ )	$-1.4 \times 10^{-8}$ ( $1.52 \times 10^{-8}$ )	
Young adult		0.00196 (0.0144)		0.00216 (0.0145)	
Mature adult		$-0.0531^{***}$ (0.0167)		$-0.0517^{***}$ (0.0167)	$-0.0528^{***}$ (0.0149)
Gestational age (weeks)			$-0.0015$ (0.00605)	$-0.0014$ (0.00605)	0.00246 (0.00436)
Male			0.0463** (0.0198)	0.0440** (0.0198)	0.0446** (0.0198)
Constant	0.705*** (0.0397)	0.716*** (0.0403)	0.711*** (0.116)	0.721*** (0.116)	0.676*** (0.105)
Observations	714	714	713	713	713
Number of individuals	287	287	286	286	286

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

**Note**

Standard errors are in brackets.

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 LR test comparing these two models [these had to be re-estimated using maximum likelihood estimation (MLE) in order to obtain the log-likelihoods necessary to calculate the LR test statistic; the parameter estimates were identical when using MLE and GLS] resulted in a  $\chi^2$  statistic of 0.9 ( $p = 0.6368$ ). Therefore, there is insufficient evidence for a quadratic relationship between birthweight and utility. As shown in (5), the marginal impact of a 150-g reduction in birthweight is a utility reduction of 0.004.

The linear birthweight coefficient was significantly different from 0 at the 10% level only in specifications in which sex and gestational age were not included. In order to test whether or not the simple specification (1) is equally as valid as (4), another LR test was conducted. The test statistic of 18.16 ( $p = 0.0011$ ) suggests that we should reject that specification (1) is equally as suitable as specification (4).

We also ran preliminary non-parametric (local polynomial) regressions that are consistent with the findings just described. These results suggest that the sample size available for analysis is too small to detect a small, if existent at all, effect of birthweight on utility, as measured by the HUI3 instrument.

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.

## Quality-adjusted life-year calculations and comparison with parameters used in the National Institute for Health and Care Excellence guidelines model

### Preterm survivors

#### National Institute for Health and Care Excellence *guidelines*

The model used for tocolytic treatment in the 2015 NICE guidelines for preterm labour<sup>24</sup> provides utilities for preterm survivors through the gestational age range of 24 to 34 weeks. The assumption made was that those surviving to the age of 1 year would live for 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) [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<sup>24</sup>); therefore, it is somewhat outdated]. 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 on data specifically from individuals born preterm.

#### Peninsula Technology Assessment Group 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 years.<sup>111</sup> The range is generated by taking the minimum and maximum mapped SF-36 values out of four other SF-36 follow-up studies of preterm individuals, assessed at a similar point in the life cycle (from 20 to 31 years of age).<sup>112,113,168,169</sup>

Rather than assuming fixed life expectancy, our model uses ONS 2014–16 life tables<sup>119</sup> for survival. These are multiplied by baseline population utilities for each age, which are derived from a regression equation by Ara and Brazier.<sup>84</sup> 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, and the Ara and Brazier<sup>84</sup> regression equation was obtained by fitting to data with an age range of 16–98 years. Finally, the utilities for each health state from *Table 20* 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<sup>186</sup> 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 on that of the 2015 NICE guidelines model<sup>24</sup> by taking population ageing and survival into account, while preserving NICE's use of survival probabilities by gestational age. Alternative QALY values were also calculated by excluding the population ageing 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.

Two studies were found using data from Sweden, which reported hazard ratios for preterm children.<sup>187,188</sup> 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.

Table 46 summarises the base-case total QALYs for preterm survivors by gestational age.

### Intraventricular haemorrhage and respiratory distress syndrome

The 2015 NICE guidelines model<sup>24</sup> for tocolytic treatment assumes a 'dummy value' for the lifetime QALY loss from RDS, owing to its variable prognosis. This value was chosen to be marginally lower than the lifetime QALY loss from IVH. The economic model in the study by Bastek *et al.*<sup>109</sup> 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 the study by Carroll and Downs<sup>108</sup> are identical at 0.91. 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 the study by Carroll and Downs<sup>108</sup> as a proxy for the lifetime effects of RDS (0.85).

**TABLE 46** Total discounted QALYs for preterm survivors, by gestational age

Gestational age (weeks)	Probability that infant survives the first year (based on ONS 2015 data) <sup>186</sup>	Weighted discounted QALYs using age and survival adjustments	Weighted discounted QALYs using NICE method (= maximum QALYs × probability of infant being alive after the first 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

N/A, not applicable.

**Note**

The NICE guidelines model<sup>24</sup> stops at 34 weeks' gestational age, and therefore no NICE QALYs are reported for the 35-week and 36-week gestational ages.



The QALY loss from IVH of 4.5 in the guidelines model was based on the value for ICH, which was assumed to be one-third of the QALY loss from moderate to severe cerebral palsy.<sup>78</sup> The source for the health-state utility on which this QALY loss was derived appears to have been obtained from Pham and Crowther<sup>174</sup> as the utility of permanent neurological sequelae, assessed by antenatal or emergency midwives.<sup>174</sup> However, a very small sample size of 14 was used to elicit this result.

In comparison, the proxy of moderate cerebral palsy used in the study by Bastek *et al.*<sup>109</sup> for IVH was obtained from the study by Carroll and Downs.<sup>108</sup> This appears to be a more reliable source, given that the time trade-off elicitation task was conducted on a much larger (and more relevant) sample of 4016 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 47* 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.

## Mothers

In common with all other previous models of preterm labour, the NICE guidelines model<sup>24</sup> does not account for the health-related quality of life of mothers of preterm children after adverse outcomes. Although data are extremely limited, we use the utility mapped from the SF-36 scores reported by Couto *et al.*<sup>106</sup> 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. 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 2015<sup>186</sup> data state that the mean age of mothers at birth is 30 years, and this is echoed by findings in *Chapter 2*. 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 47** Total discounted QALY loss for RDS and IVH as used in the economic model

Outcome	PenTAG base case	NICE guidelines <sup>24</sup> 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)

**TABLE 48** Total discounted QALYs for mothers (assuming age at birth of 30 years)

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



## Appendix 6 Additional cost-effectiveness results

### Women presenting at 26 weeks' gestation (at a level 2 hospital)

As noted previously, ICERs should be interpreted with caution because, other than 'treat all' and fFN at 10 ng/ml versus fFN at 50 ng/ml, they represent a reduction in both costs and QALYs. Incremental costs and QALYs for PartoSure versus fFN at 50 ng/ml are the result of an indirect comparison between the studies by Bruijn *et al.*<sup>42,43</sup> and Hadzi-Lega *et al.*,<sup>44</sup> because no included study directly compares these two tests. As for the case of women presenting at 30 weeks' gestation, Actim Partus results in £35,364 of cost savings per QALY lost relative to fFN at 50 ng/ml, which are higher than those of fFN at 200 ng/ml (£16,541) or fFN at 500 ng/ml (£11,476). Based on indirect comparison, PartoSure appears to offer the same QALY loss as but higher cost savings than Actim Partus, relative to fFN at 50 ng/ml.

Full incremental analyses are presented in *Table 50* and detailed costs and QALYs are presented in *Table 51*. Note that diagnostic options involving wider use of treatment have become more attractive for this group of women than for women presenting at later gestation (e.g. ICER for treat all in *Table 49*, £128,939, is lower than that in *Table 23*, £186,754).

**TABLE 49** Summary of ICERs for women presenting at 26 weeks' gestation (level 2 hospital)

Test	Total costs (£)	Total QALYs	Vs. treat all			Vs. fFN at 50 ng/ml		
			Incremental costs (£)	Incremental QALYs	ICER (£) (per QALY)	Incremental costs (£)	Incremental QALYs	ICER (£) (per QALY)
Actim Partus <sup>a</sup>	15,263	21.619	-2259	-0.031	72,794 <sup>b</sup>	-658	-0.019	35,364 <sup>b</sup>
PartoSure <sup>c</sup>	14,926 <sup>d</sup>	21.619 <sup>d</sup>	-2266	-0.024	95,252 <sup>b</sup>	-995	-0.019	53,446 <sup>b</sup>
Treat all	17,522	21.650	0	0	–	1600	0.012	128,939
fFN at 10 ng/ml <sup>a</sup>	16,498	21.643	-1024	-0.006	165,033 <sup>b</sup>	576	0.006	92,845
fFN at 50 ng/ml <sup>a</sup>	15,921	21.637	-1600	-0.012	128,939 <sup>b</sup>	0	0	–
fFN at 200 ng/ml <sup>a</sup>	15,442	21.608	-2080	-0.041	50,260 <sup>b</sup>	-479	-0.029	16,541 <sup>b</sup>
fFN at 500 ng/ml <sup>a</sup>	15,114	21.567	-2408	-0.070	29,095 <sup>b</sup>	-807	-0.070	11,476 <sup>b</sup>

a Bruijn *et al.*<sup>42,43</sup>

b The ICER represents the south-west quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

c Total costs and QALYs for PartoSure were obtained by applying the differences in these outcomes between PartoSure and Actim Partus reported in Hadzi Lega *et al.*<sup>45</sup> to corresponding results for Actim Partus reported by Bruijn *et al.*<sup>43,44</sup>

d Inferred total cost and QALYs for PartoSure obtained by applying relative differences vs. Actim Partus found using Hadzi-Lega *et al.*<sup>44</sup> to Bruijn *et al.*<sup>42,43</sup>

#### Note

Green text denotes ratios in which the difference in cost and the difference in benefit (QALYs) are both negative relative to the reference test option.

**TABLE 50** Fully incremental analysis of ICERs for women presenting at 26 weeks' gestation (level 2 hospital)

Test	Total costs (£)	Total QALYs	Vs. next option in the QALY ranking		
			Incremental costs (£)	Incremental QALYs	ICER (£)
Treat all (test none)	17,522	21.650	1024	0.006	165,033
fFN at 10 ng/ml <sup>a</sup>	16,498	21.643	576	0.006	92,845
fFN at 50 ng/ml <sup>a</sup>	15,921	21.637	658	0.019	53,446 <sup>b</sup>
Actim Partus <sup>a</sup>	15,263	21.619	337	0.000	Dominated by PartoSure
PartoSure <sup>c</sup>	14,926 <sup>d</sup>	21.619 <sup>d</sup>	-516	0.010	Dominates fFN at 200 ng/ml and fFN at 500 ng/ml
fFN at 200 ng/ml <sup>a</sup>	15,442	21.608	328	0.041	7930
fFN at 500 ng/ml <sup>a</sup>	15,114	21.567	-	-	- (dominated by PartoSure)

a Bruijn *et al.*<sup>42,43</sup>

b Calculated relative to PartoSure, which dominates Actim Partus.

c Total costs and QALYs for PartoSure were obtained by applying the differences in these outcomes between PartoSure and Actim Partus reported in Hadzi Lega *et al.*<sup>45</sup> to corresponding results for Actim Partus reported by Bruijn *et al.*<sup>43,44</sup>

d Calculated relative to PartoSure, which dominates Actim Partus.

#### 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).

**TABLE 51** Breakdown of discounted costs and QALYs for women presenting at 26 weeks' gestation

Costs and QALYs	Treat all	Bruijn <i>et al.</i> <sup>42,43</sup> (APOSTEL-1)					Indirect comparison <sup>a</sup>	
		fFN at				Actim Partus	PartoSure	
		10 ng/ml	50 ng/ml	200 ng/ml	500 ng/ml			
Discounted costs (£)								
Diagnosis	0	66	66	66	66	35	52	
Treatment	367	216	136	69	0	49	0	
Hospital admission	1325	781	493	250	95	177	1	
In-utero transfer	965	569	359	182	69	129	1	
Neonatal IVH	5232	5235	5237	5248	5264	5244	5244	
Neonatal RDS	9467	9473	9480	9509	9552	9499	9499	
Neonatal death <sup>b</sup>	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 at 50 ng/ml)	1600	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	
Infant morbidity due to IVH	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	
Infant morbidity due to RDS	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	
Infant mortality avoidance	0.13	0.13	0.12	0.10	0.06	0.10	0.10	

**TABLE 51** Breakdown of discounted costs and QALYs for women presenting at 26 weeks' gestation (*continued*)

Costs and QALYs	Treat all	Bruijn <i>et al.</i> <sup>42,43</sup> (APOSTEL-1)				Indirect comparison <sup>a</sup>	
		fFN at				Actim Partus	PartoSure
		10 ng/ml	50 ng/ml	200 ng/ml	500 ng/ml		
Total	21.650	21.643	21.637	21.608	21.567	21.619	21.619
Incremental QALYs (vs. fFN at 50 ng/ml)	0.012	0.006	Reference	-0.029	-0.070	-0.019	-0.019
ICER vs. fFN at 50 ng/ml	128,939	92,845	Reference	16,541 <sup>c</sup>	11,476 <sup>c</sup>	35,364 <sup>c</sup>	53,446 <sup>c</sup>

a Costs and QALYs are inferred values computed via an indirect comparison between the Hadzi-Lega *et al.*<sup>44</sup> and Bruijn *et al.*<sup>42,43</sup> studies.

b These are the neonatal hospital costs associated with those infants saved by steroid treatment.

c ICER represents the south-west quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

#### Note

Green text denotes ratios in which the difference in cost and the difference in benefit (QALYs) are both negative relative to the reference test option.

## Women presenting at 33 weeks' gestation (at a level 2 hospital)

Results that are qualitatively similar to those described before for women presenting at 26 and 30 weeks' gestation were found for women presenting at 33 weeks' gestation. At £97,069, Actim Partus saves more costs per QALY lost relative to fFN at 50 ng/ml than fFN at 200 ng/ml (£43,781) and fFN at 500 ng/ml (£29,631), and treat all and fFN at 10 ng/ml both have incremental costs per QALY gained that are > £200,000. Based on indirect comparison, PartoSure appears to dominate Actim Partus as it results in the same number of QALYs and lower costs. *Table 52* presents the summary results for each test relative to the comparators, *Table 53* presents the fully incremental analyses and *Table 54* presents the detailed costs and QALY elements.

**TABLE 52** Summary of ICERs for women presenting at 33 weeks' gestation (level 2 hospital)

Test	Total costs (£)	Total QALYs	Vs. treat all			Vs. fFN at 50 ng/ml		
			Incremental costs (£)	Incremental QALYs	ICER (£) (per QALY)	Incremental costs (£)	Incremental QALYs	ICER (£) (per QALY)
Actim Partus <sup>a</sup>	2716	22.096	-1117	-0.006	187,479 <sup>b</sup>	-347	-0.004	97,069 <sup>b</sup>
PartoSure <sup>c</sup>	2556 <sup>d</sup>	22.096 <sup>d</sup>	-1111	-0.005	243,269 <sup>b</sup>	-507	-0.004	141,838 <sup>b</sup>
Treat all	3833	22.102	0	0	-	770	0.002	323,093
fFN at								
10 ng/ml <sup>a</sup>	3352	22.101	-481	-0.001	403,469 <sup>b</sup>	289	0.001	242,716
50 ng/ml <sup>a</sup>	3063	22.100	-770	-0.002	323,093 <sup>b</sup>	0	0	-
200 ng/ml <sup>a</sup>	2820	22.094	-1013	-0.008	127,575 <sup>b</sup>	-243	-0.006	43,781 <sup>b</sup>
500 ng/ml <sup>a</sup>	2663	22.086	-1170	-0.016	73,650 <sup>b</sup>	-400	-0.014	29,631 <sup>b</sup>

a Bruijn *et al.*<sup>42,43</sup>

b ICER represents the south-west quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

c Total costs and QALYs for PartoSure were obtained by applying the differences in these outcomes between PartoSure and Actim Partus reported in Hadzi Lega *et al.*<sup>45</sup> to corresponding results for Actim Partus reported by Bruijn *et al.*<sup>43,44</sup>

d Inferred total cost and QALYs for PartoSure obtained by applying relative differences vs. Actim Partus found using Hadzi-Lega *et al.*<sup>44</sup> to Bruijn *et al.*<sup>42,43</sup>

#### Note

Green text denotes ratios in which the difference in cost and the difference in benefit (QALYs) are both negative relative to the reference test option.

**TABLE 53** Fully incremental analysis of ICERs for women presenting at 33 weeks' gestation (level 2 hospital)

Test	Total costs (£)	Total QALYs	Vs. next option in the QALY ranking		
			Incremental costs (£)	Incremental QALYs	ICER (£)
Treat all (test none)	3833	22.102	481	0.001	403,469
fFN at 10 ng/ml <sup>a</sup>	3352	22.101	289	0.001	242,716
fFN at 50 ng/ml <sup>a</sup>	3063	22.100	347	0.004	141,838 <sup>b</sup>
Actim Partus <sup>a</sup>	2716	22.096	160	0.000	Dominated by PartoSure
PartoSure <sup>c</sup>	2556 <sup>d</sup>	22.096 <sup>d</sup>	-264	0.002	Dominates fFN at 200 ng/ml and fFN at 500 ng/ml
fFN at 200 ng/ml <sup>a</sup>	2820	22.094	157	0.008	19,725 (dominated by PartoSure)
fFN at 500 ng/ml <sup>a</sup>	2663	22.086	-	-	- (dominated by PartoSure)

a Bruijn *et al.*<sup>42,43</sup>

b Calculated relative to PartoSure, which dominates Actim Partus.

c Total costs and QALYs for PartoSure were obtained by applying the differences in these outcomes between PartoSure and Actim Partus reported in Hadzi Lega *et al.*<sup>45</sup> to corresponding results for Actim Partus reported by Bruijn *et al.*<sup>43,44</sup>

d Inferred total cost and QALYs for PartoSure obtained by applying relative differences vs. Actim Partus found using Hadzi-Lega *et al.*<sup>44</sup> to Bruijn *et al.*<sup>42,43</sup>

#### 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).

**TABLE 54** Breakdown of discounted costs and QALYs for women presenting at 33 weeks' gestation

Costs and QALYs	Treat all	Bruijn <i>et al.</i> <sup>42,43</sup> (APOSTEL-1)				Indirect comparison <sup>a</sup>	
		fFN at				Actim Partus	PartoSure
		10 ng/ml	50 ng/ml	200 ng/ml	500 ng/ml		
Discounted costs (£)							
Diagnosis	0	66	66	66	66	35	52
Treatment	5	3	2	1	0	1	0
Hospital admission	1325	781	493	250	95	177	1
In-utero transfer	0	0	0	0	0	0	0
Neonatal IVH	2477	2478	2479	2484	2492	2482	2482
Neonatal RDS	0	0	0	0	0	0	0
Neonatal death <sup>b</sup>	26	25	24	19	11	21	21
Total	3833	3352	3063	2820	2663	2716	2556
Incremental costs (vs. fFN at 50 ng/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
Infant morbidity due to IVH	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
Infant morbidity due to RDS	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00

**TABLE 54** Breakdown of discounted costs and QALYs for women presenting at 33 weeks' gestation (*continued*)

Costs and QALYs	Treat all	Bruijn <i>et al.</i> <sup>42,43</sup> (APOSTEL-1) fFN at				Actim Partus	Indirect comparison <sup>a</sup>
		10 ng/ml	50 ng/ml	200 ng/ml	500 ng/ml		PartoSure
Loss of infant 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 at 50 ng/ml)	0.002	0.001	Reference	-0.006	-0.014	-0.004	-0.004
ICER vs. fFN at 50 ng/ml	323,093	242,716	-	43,781 <sup>c</sup>	29,631 <sup>c</sup>	97,069 <sup>c</sup>	141,838 <sup>c</sup>

a Costs and QALYs are inferred values computed via an indirect comparison between the Hadzi-Lega *et al.*<sup>44</sup> and Bruijn *et al.*<sup>42,43</sup> studies.

b These are the neonatal hospital costs associated with those infants saved by steroid treatment.

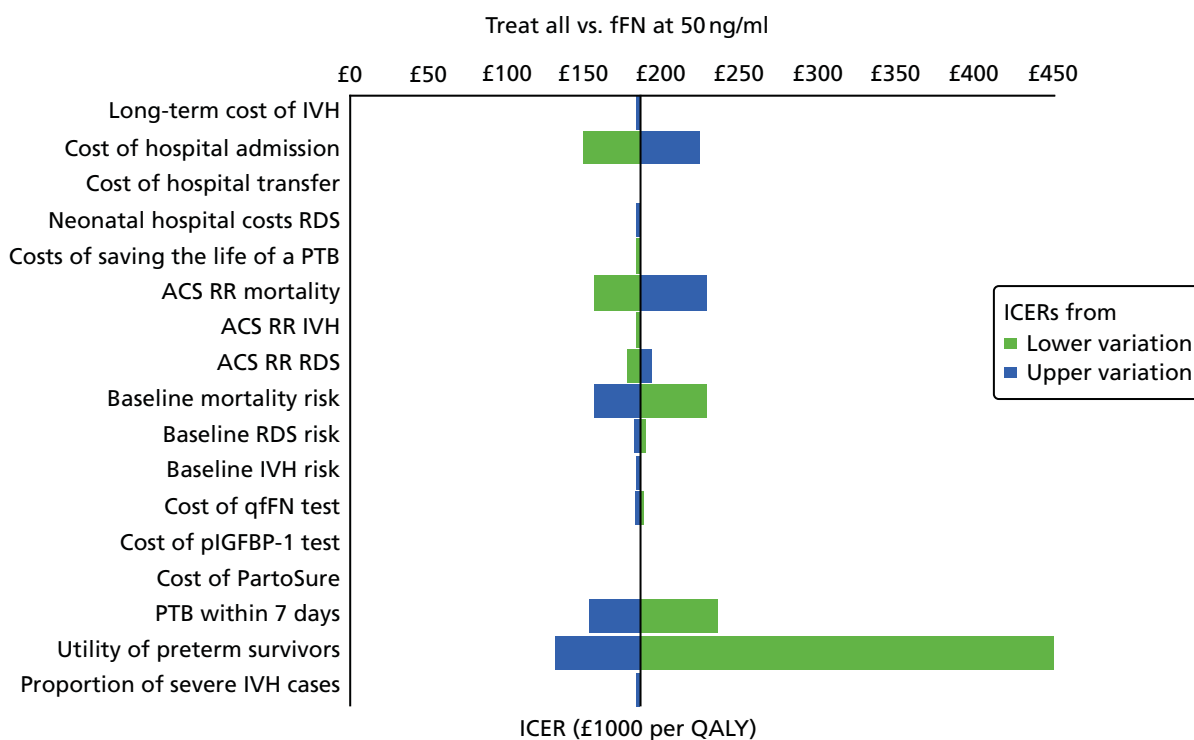
c The ICER represents the south-west quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

#### Note

Green text denotes ratios in which the difference in cost and the difference in benefit (QALYs) are both negative relative to the reference test option.

Full incremental analyses are presented in *Table 53* and detailed costs and QALYs are presented in *Table 54*. Diagnostic options involving wider use of treatment have become less attractive for this group of women than for women presenting at earlier gestation (e.g. ICER for treat all in *Table 52*, £323,093, is higher than those in *Table 49*, £128,939, and *Table 23*, £186,754).

## Additional deterministic sensitivity analyses

**FIGURE 19** Tornado diagram of most influential model parameter values: treat all. PTB, preterm birth; RR, risk ratio.

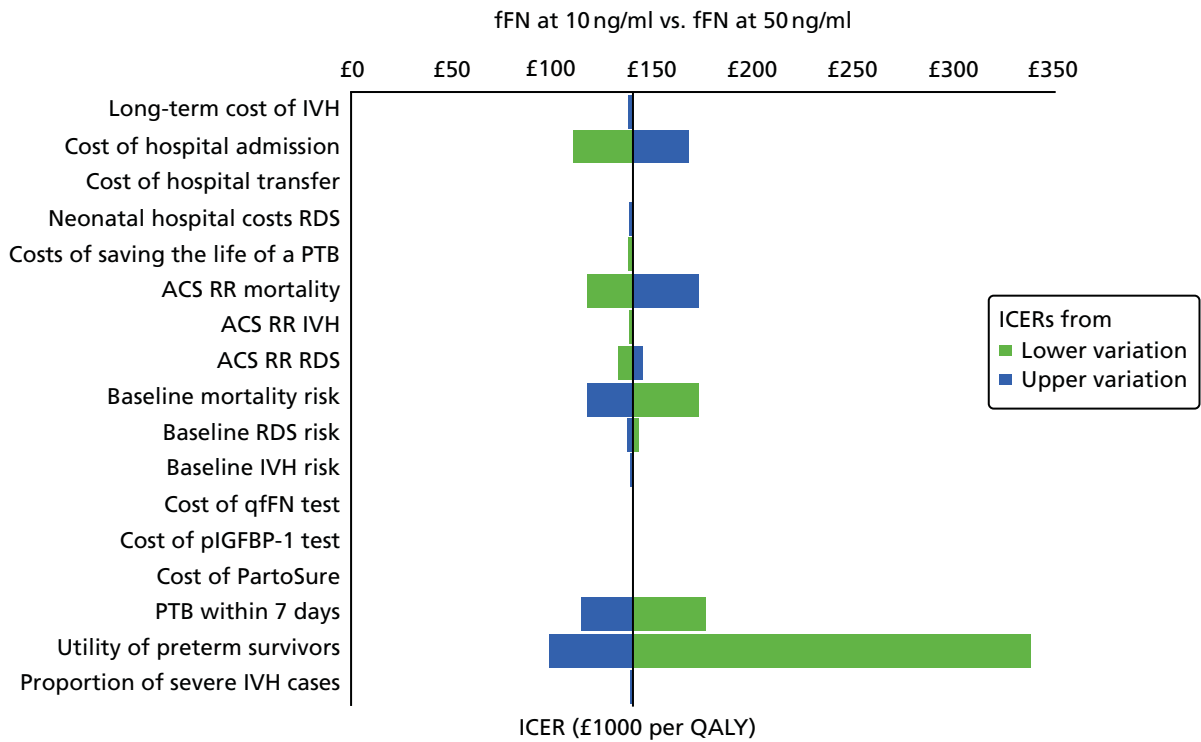


FIGURE 20 Tornado diagram of most influential model parameter values: fFN at 10 ng/ml. PTB, preterm birth; RR, risk ratio.

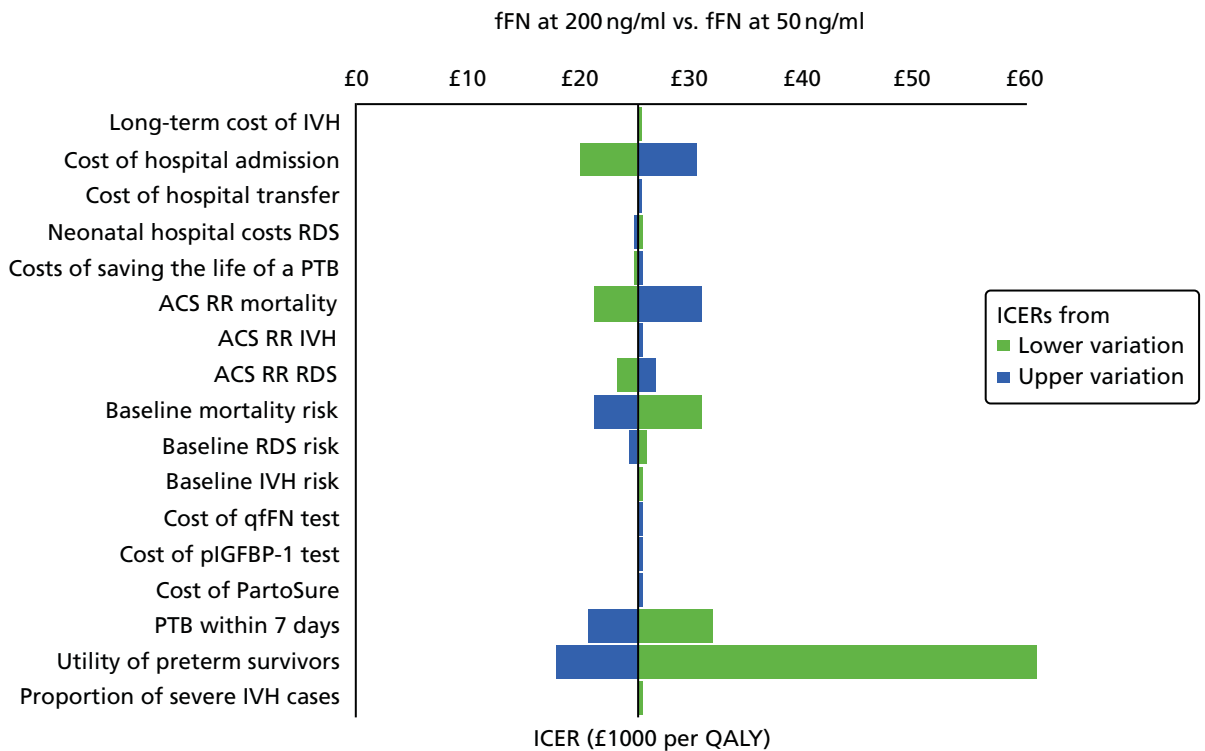
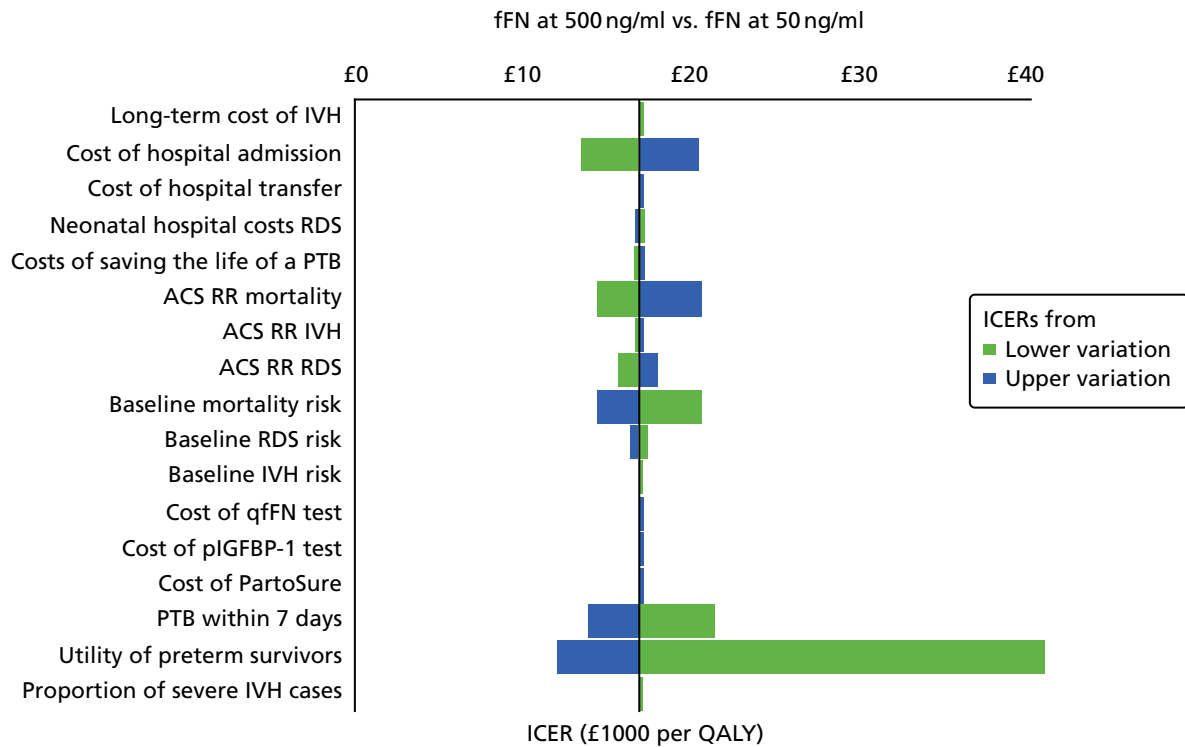
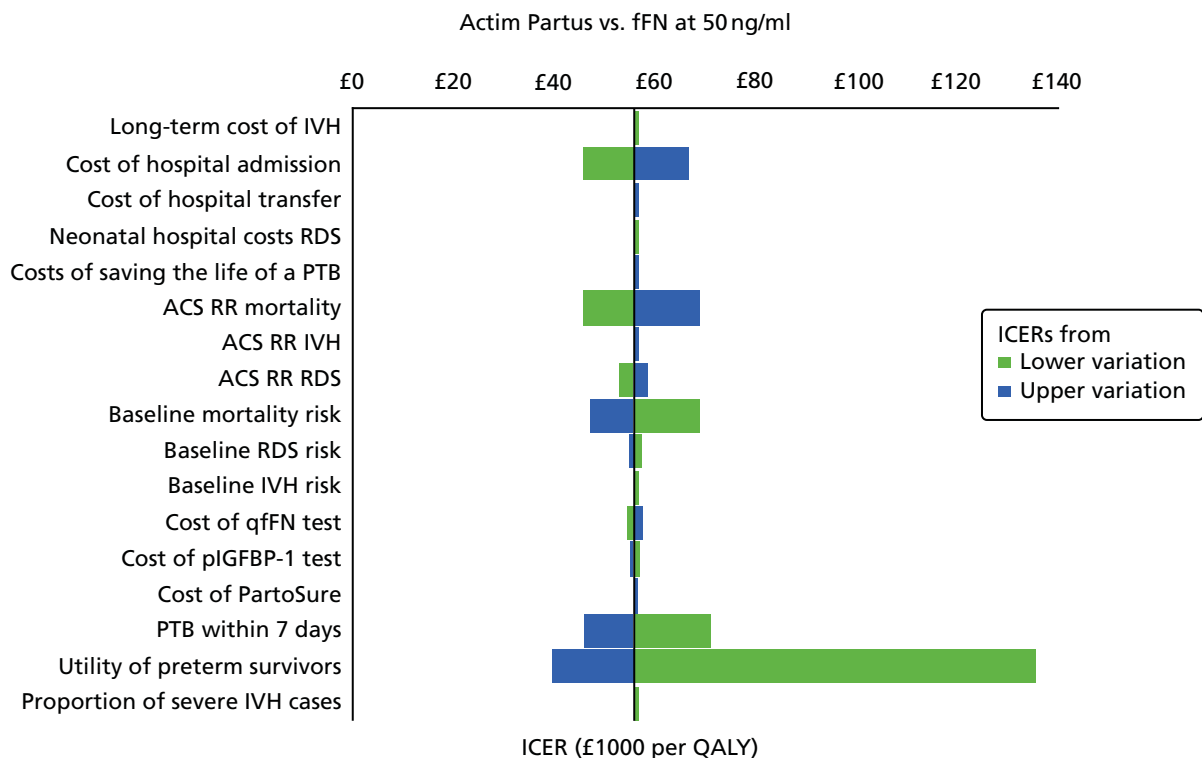


FIGURE 21 Tornado diagram of most influential model parameter values: fFN at 200 ng/ml. PTB, preterm birth; RR, risk ratio.



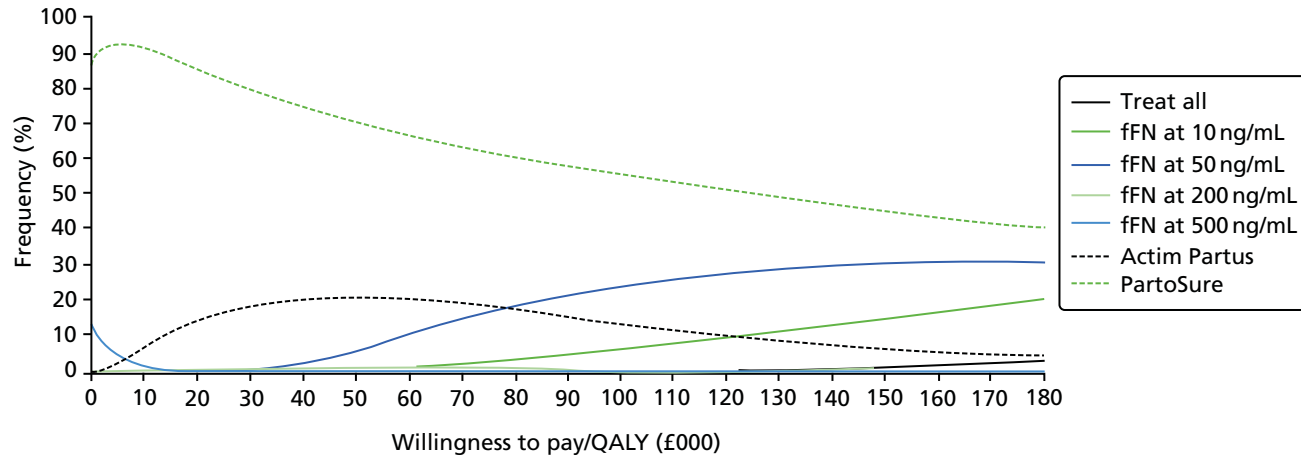


**FIGURE 22** Tornado diagram of most influential model parameter values: fFN at 500 ng/ml. PTB, preterm birth; RR, risk ratio.

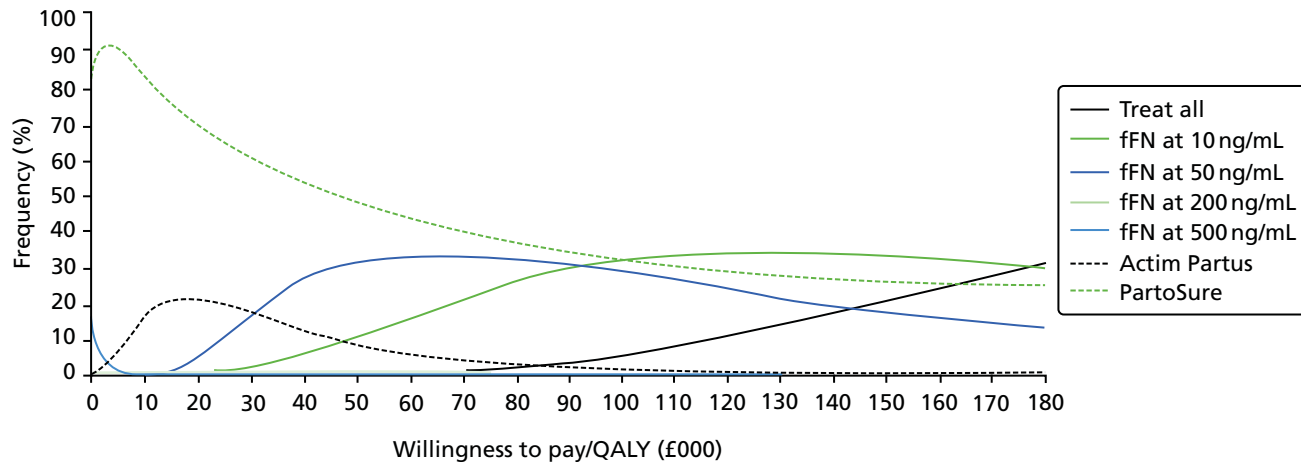


**FIGURE 23** Tornado diagram of most influential model parameter values: Actim Partus. PTB, preterm birth; RR, risk ratio.

## Additional probabilistic sensitivity analyses



**FIGURE 24** Probabilistic analysis: women presenting at 33 weeks' gestation. The highest net monetary benefit is for a gestational age at presentation of 33 weeks.



**FIGURE 25** Probabilistic analysis: women presenting at 26 weeks' gestation. The highest net monetary benefit is for a gestational age at presentation of 26 weeks.

## Scenario analyses

**TABLE 55** Results for Actim Partus and no testing vs. fFN at 50 ng/ml using data from Cooper *et al.*<sup>49</sup> for women presenting at 30 weeks' gestation (level 2 hospital)

Outcome	Cooper <i>et al.</i> <sup>49</sup>		
	Treat all	Actim Partus	fFN at 50 ng/ml
Discounted costs (£)			
Diagnosis	0	35	66
Medication	5	1	1
Admission	1325	373	171
Transfer	0	0	0
IVH	4006	4034	4034
RDS	624	630	630
Neonatal death	47	16	16
Total	6007	5090	4917
Incremental costs vs. fFN at 50 ng/ml	1090	173	
Discounted QALYs			
Baseline without morbidity	22.00	22.00	22.00
IVH	-0.02	-0.02	-0.02
RDS	-0.00	-0.00	-0.00
Infant mortality	0.04	0.01	0.01
Total	22.02	21.99	21.99
Incremental QALYs vs. fFN at 50 ng/ml	0.03	0.00	
ICER vs. fFN at 50 ng/ml	£34,508	Dominated	-

**TABLE 56** Results for fFN (various thresholds) vs. fFN at 50 ng/ml using data from Abbott *et al.*<sup>92</sup> for women presenting at 30 weeks' gestation (level 2 hospital)

Outcome	Abbott <i>et al.</i> <sup>92</sup>				
	Treat all	fFN at			
		10 ng/ml	200 ng/ml	500 ng/ml	50 ng/ml
Discounted costs					
Diagnosis	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
Medication	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
Admission	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)

continued

**TABLE 56** Results for fFN (various thresholds) vs. fFN at 50 ng/ml using data from Abbott *et al.*<sup>92</sup> for women presenting at 30 weeks' gestation (level 2 hospital) (*continued*)

Outcome	Abbott <i>et al.</i> <sup>92</sup>				
	Treat all	fFN at			
		10 ng/ml	200 ng/ml	500 ng/ml	50 ng/ml
Transfer	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
RDS	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
IVH	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
Neonatal death	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
Total	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
Incremental costs vs. fFN at 50 ng/ml	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
Discounted QALYs					
Baseline without morbidity	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
RDS	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
IVH	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
Infant mortality	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
Total	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
Incremental QALYs vs. fFN at 50 ng/ml	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)
ICER vs. fFN at 50 ng/ml	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)	(Confidential information has been removed)

**TABLE 57** Results for no testing and Actim Partus vs. fFN at 50 ng/ml using data from meta-analysis for women presenting at 30 weeks' gestation (level 2 hospital)

Outcome	Meta-analysis <sup>42,43,51,56,57</sup>		
	Treat all	Actim Partus	fFN at 50 ng/ml
Discounted costs (£)			
Diagnosis	0	35	66
Medication	5	1	1
Admission	1325	195	204
Transfer	0	0	0
IVH	4006	4013	4019
RDS	624	626	627
Neonatal death	47	39	32
Total	6007	4908	4949
Incremental costs vs. fFN at 50 ng/ml	1058	-41	
Discounted QALYs			
Baseline without morbidity	22.00	22.00	22.00
IVH	-0.02	-0.02	-0.02
RDS	-0.00	-0.00	-0.00
Infant 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 at 50 ng/ml	0.02	0.01	
ICER (relative to fFN at 50 ng/ml)	£70,468	Dominant	-

Table 58 shows results for women presenting at a gestational age of 26 weeks and Table 59 shows results for women presenting at a gestational age of 33 weeks. Results are similar to those for women presenting at 30 weeks. One main exception is that Actim Partus saves £24,532 in health-care costs per QALY lost among women with a gestational age of 33 weeks.

For women presenting at 26 weeks, excluding the neonatal hospitalisation costs associated with saving an infant's life by timely administration of ACSs has the effect of halving the ICERs relative to fFN at 50 ng/ml. Therefore, this favours the treatment-intensive options treat all and fFN at 10 ng/ml, which now have an ICER of £61,791 and £46,358, respectively. Other options are favoured by the change, but all now save < £20,000 per QALY lost relative to fFN at 50 ng/ml. The exception is PartoSure, which saves £26,988 per QALY lost.

**TABLE 58** Incremental cost-effectiveness ratios vs. fFN at 50 ng/ml for women presenting at 26 weeks' gestation (level 2 hospital)

Option	ICERs (£)							
	Base case	With maternal QALYs for 10 years	Limiting the analysis to delivery (additional cost only)	Limiting the analysis to first year after birth	ANSs earlier than 7 days before preterm delivery has partial benefits	Excluding additional neonatal hospital costs of death	Women presenting at a level 3 hospital	Applying costs and disutilities of AEs to all AEs
Treat all	128,939	72,006	1603	3,422,534	41,153	127,779	61,791	117,174
fFN at								
10 ng/ml <sup>a</sup>	92,845	45,524	578	2,470,464	23,957	91,685	46,358	84,373
200 ng/ml <sup>a</sup>	16,541 <sup>b</sup>	11,916 <sup>b</sup>	-486	457,751 <sup>b</sup>	8557 <sup>b</sup>	15,381 <sup>b</sup>	8161 <sup>b</sup>	15,032 <sup>b</sup>
500 ng/ml <sup>a</sup>	11,476 <sup>b</sup>	8660 <sup>b</sup>	-824	324,143 <sup>b</sup>	6576 <sup>b</sup>	10,316 <sup>b</sup>	5444 <sup>b</sup>	10,429 <sup>b</sup>
Actim Partus <sup>a</sup>	35,364 <sup>b</sup>	22,807 <sup>b</sup>	-663	954,254 <sup>b</sup>	14,629 <sup>b</sup>	34,204 <sup>b</sup>	18,392 <sup>b</sup>	32,137 <sup>b</sup>
PartoSure <sup>c</sup>	53,446 <sup>b</sup>	53,424 <sup>b</sup>	-1000	1,431,224 <sup>b</sup>	68,857 <sup>b</sup>	52,287 <sup>b</sup>	26,988 <sup>b</sup>	48,570 <sup>b</sup>

AE, adverse event.

a Bruijn *et al.*<sup>42,43</sup>

b The ICER represents the south-west quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

c Indirect comparison between Bruijn *et al.*<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup> (Bruijn *et al.*<sup>42,43</sup> was used as the reference study in this case).

**Note**

Green text denotes ratios in which the difference in cost and the difference in benefit (QALYs) are both negative relative to the reference test option.

**TABLE 59** Incremental cost-effectiveness ratios vs. fFN at 50 ng/ml for women presenting at 33 weeks' gestation (level 2 hospital)

Option	ICERs (£)							
	Base case	With maternal QALYs for 10 years	Limiting the analysis to delivery (additional cost only)	Limiting the analysis to first year after birth	ACs earlier than 7 days before preterm delivery has partial benefits	Excluding additional neonatal hospital costs of death	Women presenting at a 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 at								
10 ng/ml <sup>a</sup>	242,716	130,060	289	6,402,235	34,621	241,750	242,716	230,256
200 ng/ml <sup>a</sup>	43,781 <sup>b</sup>	33,081 <sup>b</sup>	-243	1,154,838 <sup>b</sup>	14,902 <sup>b</sup>	42,815 <sup>b</sup>	43,781 <sup>b</sup>	41,534 <sup>b</sup>
500 ng/ml <sup>a</sup>	29,631 <sup>b</sup>	23,314 <sup>b</sup>	-400	781,581 <sup>b</sup>	11,654 <sup>b</sup>	28,664 <sup>b</sup>	29,631 <sup>b</sup>	28,110 <sup>b</sup>
Actim Partus <sup>a</sup>	97,069 <sup>b</sup>	66,541 <sup>b</sup>	-347	2,560,443 <sup>b</sup>	24,532 <sup>b</sup>	96,103 <sup>b</sup>	97,069 <sup>b</sup>	92,086 <sup>b</sup>
PartoSure <sup>c</sup>	141,838 <sup>b</sup>	141,788 <sup>b</sup>	-507	3,741,321 <sup>b</sup>	267,481 <sup>b</sup>	140,871 <sup>b</sup>	141,838 <sup>b</sup>	134,556 <sup>b</sup>

AE, adverse event.

a Bruijn *et al.*<sup>42,43</sup>

b The ICER represents the south-west quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

c Indirect comparison between Bruijn *et al.*<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup> (Bruijn *et al.*<sup>42,43</sup> was used as the reference study in this case).

**Note**

Green text denotes ratios in which the difference in cost and the difference in benefit (QALYs) are both negative relative to the reference test option.

**TABLE 60** 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

Test	Total costs (£)	Total QALYs	Vs. treat all			Vs. fFN at 50 ng/ml		
			Incremental costs (£)	Incremental QALYs	ICER (£) (per QALY)	Incremental costs (£)	Incremental QALYs	ICER (£) (per QALY)
Actim Partus <sup>a</sup>	4891	22.016	-1116	-0.016	69,968 <sup>c</sup>	-346	-0.009	38,200 <sup>c</sup>
PartoSure <sup>b</sup>	4731 <sup>d</sup>	22.019 <sup>d</sup>	-1110	-0.013	88,385 <sup>c</sup>	-506	-0.006	81,893 <sup>c</sup>
fFN at								
10 ng/ml <sup>a</sup>	5526	22.029	-481	-0.003	159,831 <sup>c</sup>	289	0.004	74,564
50 ng/ml <sup>a</sup>	5237	22.025	-770	-0.007	111,813 <sup>c</sup>	0	0.000	-
200 ng/ml <sup>a</sup>	4995	22.012	-1012	-0.020	51,469 <sup>c</sup>	-242	-0.013	18,968 <sup>c</sup>
500 ng/ml <sup>a</sup>	4840	21.995	-1167	-0.037	31,829 <sup>c</sup>	-398	-0.030	13,347 <sup>c</sup>

a Bruijn *et al.*<sup>42,43</sup>

b Hadzi-Lega *et al.*<sup>44</sup> for comparison with treat all, indirect comparison between Bruijn *et al.*<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup> for comparison with fFN at 50 ng/ml (Bruijn *et al.*<sup>42,43</sup> was used as the reference study in this case).

c The ICER represents the south-west quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

d Total costs and QALYs for PartoSure were obtained by applying the differences in these outcomes between PartoSure and Actim Partus reported in Hadzi Lega *et al.*<sup>45</sup> to corresponding results for Actim Partus reported by Bruijn *et al.*<sup>43,44</sup>

**Note**

Green text denotes ratios in which the difference in cost and the difference in benefit (QALYs) are both negative relative to the reference test option.

**TABLE 61** 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

Test	Total costs (£)	Total QALYs	Vs. treat all			Vs. fFN at 50 ng/ml		
			Incremental costs (£)	Incremental QALYs	ICER (£) (per QALY)	Incremental costs (£)	Incremental QALYs	ICER (£) (per QALY)
Actim Partus <sup>a</sup>	4891	22.027	-1116	-0.025	43,954 <sup>d</sup>	-346	-0.014	24,936 <sup>d</sup>
PartoSure <sup>b</sup>	4731 <sup>c</sup>	22.035 <sup>c</sup>	-1110	-0.020	54,536 <sup>d</sup>	-506	-0.006	81,844 <sup>d</sup>
fFN at								
10 ng/ml <sup>a</sup>	5526	22.048	-481	-0.005	104,731 <sup>d</sup>	289	0.007	41,820
50 ng/ml <sup>a</sup>	5237	22.041	-770	-0.011	66,923 <sup>d</sup>	0	0.000	-
200 ng/ml <sup>a</sup>	4995	22.023	-1012	-0.030	34,226 <sup>d</sup>	-242	-0.018	13,416 <sup>d</sup>
500 ng/ml <sup>a</sup>	4840	22.000	-1167	-0.052	22,426 <sup>d</sup>	-398	-0.041	9806 <sup>d</sup>

a Bruijn *et al.*<sup>42,43</sup>

b Hadzi-Lega *et al.*<sup>44</sup> for comparison with treat all, indirect comparison between Bruijn *et al.*<sup>42,43</sup> and Hadzi-Lega *et al.*<sup>44</sup> for comparison with fFN at 50 ng/ml (Bruijn *et al.*<sup>42,43</sup> was used as the reference study in this case).

c Total costs and QALYs for PartoSure were obtained by applying the differences in these outcomes between PartoSure and Actim Partus reported in Hadzi Lega *et al.*<sup>45</sup> to corresponding results for Actim Partus reported by Bruijn *et al.*<sup>43,44</sup>

d The ICER represents the south-west quadrant in cost-effectiveness (i.e. a reduction in both costs and QALYs).

**Note**

Green text denotes ratios in which the difference in cost and the difference in benefit (QALYs) are both negative relative to the reference test option.



# Appendix 7 Additional diagnostic test accuracy data on cervical length

## Methods of the overview

In addition to the systematic review described in *Chapter 2, Methods of the systematic review*, an overview of studies that assess the DTA of at least one of the index tests (PartoSure, Actim Partus and qfFN), in addition to a qualitative fFN test, qfFN at 50 ng/ml and/or transvaginal cervical length measurement, was provided because these are the tests currently recommended by NICE guidance.<sup>24</sup>

These data were extracted, when available, from the studies included in the systematic review of PartoSure, Actim Partus and qfFN (see *Chapter 2, Results of the systematic review*). For this reason, studies included in the systematic review that also provided test accuracy data for a qualitative fFN test, qfFN at 50 ng/ml and/or transvaginal cervical length measurement, were included in this overview. Test accuracy data for qualitative fFN, qfFN at 50 ng/ml and/or transvaginal cervical length measurement were extracted, tabulated and analysed following the same methods and principles described in *Chapter 2, Methods of the systematic review*, 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 qfFN, 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, qfFN at 50 ng/ml and/or transvaginal cervical length measurement), this summary does not systematically cover the full breadth of DTA evidence on these additional tests. [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 appendix.] In order to ensure that the test accuracy data for qualitative fFN, qfFN at 50 ng/ml and/or transvaginal cervical length measurement 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.

## Results of the overview: fetal fibronectin

In this section, we present an overview of studies that assess the DTA of at least one of the index tests (PartoSure, Actim Partus and qfFN), in addition to a qualitative fFN test and/or qfFN at 50-ng/ml test. In a similar manner, a further overview incorporating cervical length is given in *Appendix 2*.

The reasons for presenting these additional data are twofold. First, fFN at 50 ng/ml is recommended in the NICE guidelines for current practice.<sup>24</sup> 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 50 ng/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. Owing to this, these data are not exhaustive of all data available for qualitative fFN or qfFN at 50 ng/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 qfFN at 50 ng/ml.

In this overview, DTA data, against a reference standard of preterm birth within 7 days, are provided for fFN, at a threshold of 50 ng/ml (qualitative or quantitative test). These data are presented in *Chapter 2, Results of the systematic review*, 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 qfFN at thresholds other than 50 ng/ml).

**TABLE 62** Test accuracy results (against preterm birth within 7 days) for index tests, cervical length measurement and fFN at 50 ng/ml

Study (study name/first author and year) and test	Number of participants	Test accuracy results (%) (95% CI)			
		Sensitivity	Specificity	PPV	NPV
Azlin (2010) <sup>47</sup>					
CL < 25 mm	51	80.0 (28.4 to 99.5)	71.7 (56.5 to 84.0)	23.5 (6.8 to 49.9)	97.1 (84.7 to 99.9)
Actim Partus	51	80.0 (28.4 to 99.5)	93.5 (82.1 to 98.6)	57.1 (18.4 to 90.1)	97.7 (88.0 to 99.9)
Bolotskikh (2017) <sup>59</sup>					
CL < 15 mm <sup>a</sup>	99	33.3 (9.9 to 65.1)	98.9 (93.8 to 100.0)	80.0 (28.4 to 99.5)	91.5 (83.9 to 96.3)
PartoSure	99	100.0 (73.5 to 100.0)	95.4 (88.6 to 98.7)	75.0 (47.6 to 92.7)	100.0 (95.7 to 100.0)
APOSTEL-1 (2016) <sup>42,43</sup>					
CL < 15 mm <sup>a</sup>	350	72.5 (60.4 to 82.5)	83.3 (78.4 to 87.4)	51.5 (41.2 to 61.8)	92.5 (88.5 to 95.4)
fFN at 10 ng/ml	350	95.7 (87.8 to 99.1)	42.3 (36.5 to 48.4)	28.9 (23.2 to 35.3)	97.5 (93.0 to 99.5)
fFN at 50 ng/ml <sup>b</sup>	350	91.3 (82.0 to 96.7)	64.8 (58.9 to 70.3)	38.9 (31.3 to 46.9)	96.8 (93.2 to 98.8)
fFN at 200 ng/ml	350	71.0 (58.8 to 81.3)	83.6 (78.8 to 87.8)	51.6 (41.1 to 62.0)	92.2 (88.1 to 95.1)
fFN at 500 ng/ml	350	42.0 (30.2 to 54.5)	95.7 (92.7 to 97.8)	70.7 (54.5 to 83.9)	87.1 (82.8 to 90.6)
Actim Partus	350	78.3 (66.7 to 87.3)	89.3 (85.1 to 92.7)	64.3 (53.1 to 74.4)	94.4 (90.9 to 96.8)
Bruijn (2016) <sup>62</sup>					
CL < 15 mm <sup>a</sup>	450	51.3 (34.8 to 67.6)	81.8 (77.7 to 85.4)	21.3 (13.5 to 30.9)	94.6 (91.7 to 96.7)
fFN at 10 ng/ml	455	93.8 (82.8 to 98.7)	32.2 (27.7 to 37.0)	14.0 (10.4 to 18.3)	97.8 (93.6 to 99.5)
fFN at 50 ng/ml <sup>b</sup>	455	89.6 (77.3 to 96.5)	62.2 (57.3 to 66.9)	21.8 (16.3 to 28.3)	98.1 (95.5 to 99.4)
fFN at 200 ng/ml	455	70.8 (55.9 to 83.0)	78.6 (74.3 to 82.5)	28.1 (20.3 to 37.0)	95.8 (93.1 to 97.7)
fFN at 500 ng/ml	455	29.2 (17.0 to 44.1)	94.3 (91.6 to 96.4)	37.8 (22.5 to 55.2)	91.9 (88.8 to 94.3)
Cooper (2012) <sup>49</sup>					
fFN at 50 ng/ml <sup>c</sup>	291	33.3 (4.3 to 77.7)	89.8 (85.7 to 93.1)	6.5 (0.8 to 21.4)	98.5 (96.1 to 99.6)
Actim Partus	349	33.3 (4.3 to 77.7)	74.1 (69.1 to 78.6)	2.2 (0.3 to 7.7)	98.4 (96.1 to 99.6)
Danti (2011) <sup>50</sup>					
CL < 20 mm <sup>a</sup> (sample 1)	60	75.0 (19.4 to 99.4)	71.4 (57.8 to 82.7)	15.8 (3.4 to 39.6)	97.6 (87.1 to 99.9)
CL < 20 mm <sup>a</sup> (sample 2)	102	75.0 (19.4 to 99.4)	83.7 (74.8 to 90.4)	15.8 (3.4 to 39.6)	98.8 (93.5 to 100.0)
Actim Partus	60	50.0 (6.8 to 93.2)	69.6 (55.9 to 81.2)	10.5 (1.3 to 33.1)	95.1 (83.5 to 99.4)
Eroglu (2007) <sup>51</sup>					
fFN at 50 ng/ml <sup>d</sup>	51	83.3 (35.9 to 99.6)	80.0 (65.4 to 90.4)	35.7 (12.8 to 64.9)	97.3 (85.8 to 99.9)
Actim Partus	51	83.3 (35.9 to 99.6)	84.4 (70.5 to 93.5)	41.7 (15.2 to 72.3)	97.4 (86.5 to 99.9)
CL < 20 mm	51	66.7 (22.3 to 95.7)	95.6 (84.9 to 99.5)	66.7 (22.3 to 95.7)	95.6 (84.9 to 99.5)
Goyal (2016) <sup>52</sup>					
CL < 25 mm <sup>e</sup>	60	80.5 (65.1 to 91.2)	31.6 (12.6 to 56.6)	71.7 (56.5 to 84.0)	42.9 (17.7 to 71.1)
Actim Partus	60	59.1 (43.2 to 73.7)	50.0 (24.7 to 75.3)	76.5 (58.8 to 89.3)	30.8 (14.3 to 51.8)

**TABLE 62** Test accuracy results (against preterm birth within 7 days) for index tests, cervical length measurement and fFN at 50 ng/ml (*continued*)

Study (study name/first author and year) and test	Number of participants	Test accuracy results (%) (95% CI)			
		Sensitivity	Specificity	PPV	NPV
Hadzi-Lega (2017) <sup>44</sup>					
CL < 25 mm	57	100.0 (54.1 to 100.0)	70.6 (56.2 to 82.5)	28.6 (11.3 to 52.2)	100.0 (90.3 to 100.0)
Actim Partus	57	83.3 (35.9 to 99.6)	76.5 (62.5 to 87.2)	29.4 (10.3 to 56.0)	97.5 (86.8 to 99.9)
PartoSure	57	83.3 (35.9 to 99.6)	90.2 (78.6 to 96.7)	50.0 (18.7 to 81.3)	97.9 (88.7 to 99.9)
Nikolova (2014 and 2015) <sup>60,61</sup>					
CL < 25 mm	203	57.1 (39.4 to 73.7)	72.6 (65.2 to 79.2)	30.3 (19.6 to 42.9)	89.1 (82.6 to 93.7)
fFN at 50 ng/ml <sup>d</sup>	66	50.0 (21.1 to 78.9)	72.2 (58.4 to 83.5)	28.6 (11.3 to 52.2)	86.7 (73.2 to 94.9)
PartoSure	203	80.0 (63.1 to 91.6)	94.6 (90.1 to 97.5)	75.7 (58.8 to 88.2)	95.8 (91.5 to 98.3)
Riboni (2011) <sup>54</sup>					
fFN at 50 ng/ml <sup>f</sup>	210	50.0 (15.7 to 84.3)	80.2 (74.0 to 85.5)	9.1 (2.5 to 21.7)	97.6 (93.9 to 99.3)
Actim Partus	210	50.0 (15.7 to 84.3)	83.7 (77.8 to 88.5)	10.8 (3.0 to 25.4)	97.7 (94.2 to 99.4)
Ting (2007) <sup>56</sup>					
fFN at 50 ng/ml <sup>c</sup>	94	56.3 (29.9 to 80.2)	75.6 (64.6 to 84.7)	32.1 (15.9 to 52.4)	89.4 (79.4 to 95.6)
Actim Partus	94	70.6 (44.0 to 89.7)	77.9 (67.0 to 86.6)	41.4 (23.5 to 61.1)	92.3 (83.0 to 97.5)
Tripathi (2016) <sup>57</sup>					
fFN at 50 ng/ml <sup>d</sup>	468	23.8 (17.3 to 31.4)	99.1 (97.3 to 99.8)	92.3 (79.1 to 98.4)	73.2 (68.7 to 77.3)
Actim Partus	467	94.7 (89.9 to 97.7)	92.4 (88.9 to 95.1)	85.7 (79.5 to 90.6)	97.3 (94.8 to 98.8)

CL, cervical length.

a Other cut-off points available.

b Quantitative Rapid fFN 10Q Cassette.

c fFN testing method unclear.

d QuikCheck fFN.

e Raw data back calculated from sensitivity, specificity, PPV and NPV.

f fFN measured by ELISA.

### Test accuracy data for fetal fibronectin at 50 ng/ml

#### Quantity and quality of the data available for fetal fibronectin at 50 ng/ml

As can be seen in *Table 63*, eight of the 20 included studies<sup>42,43,49,51,54,56,57,60–62</sup> report DTA data for fFN measured at a 50-ng/ml threshold (in addition to data for at least one index test).

Two studies (APOSTEL-1<sup>42,43</sup> and Bruijn<sup>62</sup>) used the qfFN test and report data at 50 ng/ml. The APOSTEL-1 study did additionally use a qualitative version of the fFN test (Rapid fFN for TLI<sub>IQ</sub> system test), but data for this test are not provided in the included papers.<sup>42,43</sup> Three studies<sup>51,57,60,61</sup> used the QuikCheck version of the qualitative fFN test. One further study<sup>54</sup> used the ELISA laboratory technique. The remaining two studies<sup>49,56</sup> did not report which test was used. More specifically, Ting *et al.*<sup>56</sup> only state that the test used was a bedside test that was 'qualitatively reported'. Cooper *et al.*<sup>49</sup> report using a fFN test manufactured by Adeza Biochemical Corporation, but as this company produces both an ELISA testing method and the Rapid fFN for TLI<sub>IQ</sub> system test,<sup>189</sup> it remains unclear which test was used.

**TABLE 63** Test accuracy results (against preterm birth within 7 days) for index tests and fFN at 50 ng/ml

Study (study name/first author and year) and test	Number of participants	Test accuracy results (%) (95% CI)			
		Sensitivity	Specificity	PPV	NPV
APOSTEL-1 (2016) <sup>42,43</sup>					
fFN at 10 ng/ml	350	95.7 (87.8 to 99.1)	42.3 (36.5 to 48.4)	28.9 (23.2 to 35.3)	97.5 (93.0 to 99.5)
fFN at 50 ng/ml <sup>a</sup>	350	91.3 (82.0 to 96.7)	64.8 (58.9 to 70.3)	38.9 (31.3 to 46.9)	96.8 (93.2 to 98.8)
fFN at 200 ng/ml	350	71.0 (58.8 to 81.3)	83.6 (78.8 to 87.8)	51.6 (41.1 to 62.0)	92.2 (88.1 to 95.1)
fFN at 500 ng/ml	350	42.0 (30.2 to 54.5)	95.7 (92.7 to 97.8)	70.7 (54.5 to 83.9)	87.1 (82.8 to 90.6)
Actim Partus	350	78.3 (66.7 to 87.3)	89.3 (85.1 to 92.7)	64.3 (53.1 to 74.4)	94.4 (90.9 to 96.8)
Bruijn (2016) <sup>62</sup>					
fFN at 10 ng/ml	455	93.8 (82.8 to 98.7)	32.2 (27.7 to 37.0)	14.0 (10.4 to 18.3)	97.8 (93.6 to 99.5)
fFN at 50 ng/ml <sup>a</sup>	455	89.6 (77.3 to 96.5)	62.2 (57.3 to 66.9)	21.8 (16.3 to 28.3)	98.1 (95.5 to 99.4)
fFN at 200 ng/ml	455	70.8 (55.9 to 83.0)	78.6 (74.3 to 82.5)	28.1 (20.3 to 37.0)	95.8 (93.1 to 97.7)
fFN at 500 ng/ml	455	29.2 (17.0 to 44.1)	94.3 (91.6 to 96.4)	37.8 (22.5 to 55.2)	91.9 (88.8 to 94.3)
Cooper (2012) <sup>49</sup>					
fFN at 50 ng/ml <sup>b</sup>	291	33.3 (4.3 to 77.7)	89.8 (85.7 to 93.1)	6.5 (0.8 to 21.4)	98.5 (96.1 to 99.6)
Actim Partus	349	33.3 (4.3 to 77.7)	74.1 (69.1 to 78.6)	2.2 (0.3 to 7.7)	98.4 (96.1 to 99.6)
Eroglu (2007) <sup>51</sup>					
fFN at 50 ng/ml <sup>c</sup>	51	83.3 (35.9 to 99.6)	80.0 (65.4 to 90.4)	35.7 (12.8 to 64.9)	97.3 (85.8 to 99.9)
Actim Partus	51	83.3 (35.9 to 99.6)	84.4 (70.5 to 93.5)	41.7 (15.2 to 72.3)	97.4 (86.5 to 99.9)
Nikolova (2014 and 2015) <sup>60,61</sup>					
fFN at 50 ng/ml <sup>c</sup>	66	50.0 (21.1 to 78.9)	72.2 (58.4 to 83.5)	28.6 (11.3 to 52.2)	86.7 (73.2 to 94.9)
PartoSure	203	80.0 (63.1 to 91.6)	94.6 (90.1 to 97.5)	75.7 (58.8 to 88.2)	95.8 (91.5 to 98.3)
Riboni (2011) <sup>54</sup>					
fFN at 50 ng/ml <sup>d</sup>	210	50.0 (15.7 to 84.3)	80.2 (74.0 to 85.5)	9.1 (2.5 to 21.7)	97.6 (93.9 to 99.3)
Actim Partus	210	50.0 (15.7 to 84.3)	83.7 (77.8 to 88.5)	10.8 (3.0 to 25.4)	97.7 (94.2 to 99.4)
Ting (2007) <sup>56</sup>					
fFN at 50 ng/ml <sup>b</sup>	94	56.3 (29.9 to 80.2)	75.6 (64.6 to 84.7)	32.1 (15.9 to 52.4)	89.4 (79.4 to 95.6)
Actim Partus	94	70.6 (44.0 to 89.7)	77.9 (67.0 to 86.6)	41.4 (23.5 to 61.1)	92.3 (83.0 to 97.5)
Tripathi (2016) <sup>57</sup>					
fFN at 50 ng/ml <sup>c</sup>	468	23.8 (17.3 to 31.4)	99.1 (97.3 to 99.8)	92.3 (79.1 to 98.4)	73.2 (68.7 to 77.3)
Actim Partus	467	94.7 (89.9 to 97.7)	92.4 (88.9 to 95.1)	85.7 (79.5 to 90.6)	97.3 (94.8 to 98.8)

a Quantitative Rapid fFN 10Q Cassette.

b fFN testing method unclear.

c QuikCheck fFN.

d fFN measured by ELISA.

As previously mentioned, in the APOSTEL-1 study, samples for the qfFN test (at 50 ng/ml) were collected and frozen for later analysis (this was not the case for the Rapid fFN for TLi<sub>Q</sub> 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.<sup>42,43</sup> However, as previously noted, owing to the nature of the test, the potential for interpretation bias is minimal. For all other studies reporting fFN data at 50 ng/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,<sup>51,54,57</sup> it is unclear whether or not assessors were aware of the Actim Partus test results when analysing fFN tests. 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 *et al.*<sup>54</sup>) uses the ELISA technique to determine fFN status (and possibly Cooper *et al.*<sup>49</sup> as well, although this is unclear). ELISA is a quantitative technique that was used in a qualitative capacity using 50 ng/ml as the threshold; this is the standard threshold, suggesting that in this study the threshold was established a priori.<sup>54</sup> However, neither Riboni *et al.*<sup>54</sup> nor Cooper *et al.*<sup>49</sup> explicitly report prespecification of the threshold for this test. Of course, the APOSTEL-1 study also uses a qfFN test, but multiple prespecified thresholds were used.<sup>42,43</sup>

An additional consideration in these 50-ng/ml fFN data is that one study<sup>60,61</sup> reports fFN at 50-ng/ml accuracy data for only 66 out of the 203 patients recruited and included in the PartoSure analyses. The reasons for this are unclear.

### Test accuracy of 50-ng/ml threshold for fetal fibronectin

Diagnostic test accuracy data for fFN at a threshold of 50 ng/ml (against the 7-day delivery reference standard) are provided in *Table 63*. Sensitivity of fFN at the 50-ng/ml threshold ranged from 23.8% (95% CI 17.3% to 31.4%) in the study by Tripathi *et al.*<sup>57</sup> to 91.3% (95% CI 82.0% to 96.7%) in APOSTEL-1<sup>42,43</sup> (qfFN data at 50 ng/ml). Specificity ranged from 62.2% (95% CI 57.3% to 66.9%) in Bruijn<sup>62</sup> to 99.1% (95% CI 97.3% to 99.8%) in Tripathi *et al.*<sup>57</sup> Values for PPV and NPV were also calculated and are presented in *Table 63*.

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

### Comparison of fetal fibronectin at 50-ng/ml test and index tests

In six studies,<sup>42,43,49,51,54,56,57</sup> both fFN at a threshold of 50 ng/ml and Actim Partus were assessed in the same sample. One study<sup>60,61</sup> assessed fFN at a threshold of 50 ng/ml and PartoSure in the same sample, and two studies<sup>42,43,62</sup> assessed fFN at a threshold of 50 ng/ml and fFN at other thresholds in the same sample. Note that the APOSTEL-1 study assessed more than one index test, in addition to fFN at 50 ng/ml, in the same sample.<sup>42,43</sup>

When compared with Actim Partus, sensitivity (against the reference standard of preterm birth within 7 days) was higher for fFN at 50 ng/ml in one study,<sup>42,43</sup> lower for fFN at 50 ng/ml in two studies<sup>56,57</sup> and the same for both tests in three studies.<sup>49,51,54</sup> Specificity (against the reference standard of preterm birth within 7 days) was lower for fFN than for Actim Partus in four of the six studies,<sup>42,43,51,54,56</sup> and higher for the other two studies.<sup>49,57</sup> These data are presented in *Table 63*.

In the study that included both PartoSure and fFN at a threshold of 50 ng/ml,<sup>60,61</sup> both sensitivity and specificity (against the reference standard of preterm birth within 7 days) were higher for PartoSure [sensitivity 80% (95% CI 63.1% to 91.6%) and specificity 94.6% (95% CI 90.1% to 97.5%)] than for fFN at 50 ng/ml [sensitivity 50.0% (95% CI 21.1% to 78.9%) and specificity 72.2% (95% CI 58.4% to 83.5%)].

As would be expected, in the two studies assessing qfFN at a variety of thresholds (APOSTEL-1 and Bruijn), as the threshold of fFN increased (< 10, < 50, < 200 or < 500 ng/ml) sensitivity decreased and the specificity increased.<sup>42,43,62</sup>

### Relevant systematic reviews

The test accuracy data for fFN at 50 ng/ml presented above are based only on the studies included in the systematic review of PartoSure, Actim Partus and qfFN (see *Chapter 2, Results of the systematic review*). Recent systematic reviews were sought in order to identify other available data on fFN at 50 ng/ml (either the older qualitative test or the modern quantitative test) in the prediction of preterm birth.

#### Data from systematic reviews of fetal fibronectin at a threshold of 50 ng/ml

One notable systematic review by Sanchez-Ramos *et al.*<sup>128</sup> (2009), included 32 studies that used either a qualitative fFN test or the quantitative test at a threshold of 50 ng/ml. In this review, pooled sensitivity (against a reference standard of delivery within 7 days) was 76.1% (95% CI 69.1% to 81.9%) and pooled specificity was 81.9% (95% CI 78.9% to 84.5%).<sup>128</sup> More recently, Boots *et al.*<sup>129</sup> (2014) published a systematic review that included both studies assessing fFN at a threshold of 50 ng/ml and studies assessing cervical length measurement. For fFN at 50 ng/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 *et al.*;<sup>128</sup> in the Boots *et al.*<sup>129</sup> review, pooled sensitivity was 75% (95% CI 69% to 80%) and pooled specificity was 79% (95% CI 76% to 83%). These values are also similar to those reported in recent NICE guidance,<sup>24</sup> in which 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 ranged from 61.9% (95% CI 59.6% to 62.5%) to 92% (95% CI not reported).

These systematic review data are also similar to the data for fFN at a threshold of 50 ng/ml from the current overview (see *Chapter 2, Heterogeneity between studies*), which also ranged widely: sensitivity (against the 7-day reference standard) ranged from 23.8% (95% CI 17.3% to 31.4%) to 91.3% (95% CI 82.0% to 96.7%) and specificity ranged from 62.2% (95% CI 57.3% to 66.9%) to 99.1% (95% CI 97.3% to 99.8%).

### Test accuracy data for transvaginal cervical length

#### Quantity and quality of the data available for cervical length

As can be seen in *Table 62*, nine of the 20 included studies<sup>42–44,47,50–52,59–62</sup> report DTA data for cervical length (in addition to data for at least one index test).

Of these nine studies, four<sup>42,43,51,59,62</sup> used the cervical length threshold recommended in the current NICE guidance (< 15 mm).<sup>24</sup> Three of these studies<sup>42,43,59,62</sup> also reported test accuracy data at other thresholds (not presented in this report). One study<sup>50</sup> reported test accuracy of cervical length using the threshold of < 20 mm. The remaining four studies<sup>44,47,52,60,61</sup> all used a cervical length threshold of < 25 mm.

One study<sup>50</sup> reports cervical length test accuracy data for two populations; all recruited women had their cervical length measured. For those women with a cervical length of > 30 mm, the Actim Partus test was not conducted ( $n = 42$ ), and for those with a cervical length of < 30 mm, Actim Partus was conducted ( $n = 60$ ). Cervical length test accuracy data were available both for the women with a cervical length of < 30 mm ( $n = 60$ ) (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, qfFN 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 or not the clinicians measuring cervical length were blinded to the results of any biomedical test used. Indeed, with the exception of the study by Eroglu *et al.*,<sup>51</sup> all studies that evaluated cervical length did not clearly describe whether or not clinicians were blinded to other test results.<sup>42–44,47,50,52,59–62</sup> In Eroglu *et al.*,<sup>51</sup> it was explicitly stated that the assessor was blinded to other test results.

### Test accuracy of transvaginal cervical length measurement

Table 62 provides DTA data for the three studies<sup>42,43,59,62</sup> assessing cervical length at a threshold of < 15 mm (against the 7-day delivery reference standard). At this threshold, sensitivity ranged widely, from 33.3% (95% CI 9.9% to 65.1%) in Bolotskikh *et al.*<sup>59</sup> to 72.5% (95% CI 60.4% to 82.5%) in APOSTEL-1.<sup>42,43</sup> Specificity was more similar across the four studies, ranging from 81.8% (95% CI 77.7% to 85.4%) in Bruijn<sup>62</sup> to 98.9% (95% CI 93.8% to 100.0%) in Bolotskikh *et al.*<sup>59</sup>

Against the 7-day delivery reference standard, sensitivity of cervical length at a threshold of < 20 mm was 75.0% (95% CI 19.4% to 99.4%) in both of the Danti *et al.*<sup>50</sup> samples ( $n = 60$  and  $n = 102$ ), whereas specificity was 83.7% (95% CI 74.8% to 90.4%) for the sample of  $n = 102$  and 71.4% (95% CI 57.8% to 82.7%) in the sample of  $n = 60$  (i.e. women with a cervical length of < 30 mm). Eroglu *et al.*<sup>51</sup> also evaluated cervical length at a threshold of < 20 mm against the 7-day delivery reference standard; sensitivity was lower at 66.7% (95% CI 22.3% to 95.7%), but specificity was higher at 95.6% (95% CI 84.9% to 99.5%) than Danti *et al.*<sup>50</sup> Across the four studies providing data at a threshold of < 25 mm, and again against a 7-day delivery reference standard, sensitivity ranged from 57.1% (95% CI 39.4% to 73.7%) in Nikolova *et al.*<sup>60,61</sup> to 100.0% (95% CI 54.1% to 100.0%) in Hadzi-Lega *et al.*<sup>44</sup> In these four studies, specificity ranged from 31.6% (95% CI 12.6% to 56.6%) in Goyal *et al.*<sup>52</sup> to 72.6% (95% CI 65.2% to 79.2%) in Nikolova *et al.*<sup>60,61</sup>

Again, it should be noted that these data do not cover all available evidence regarding test accuracy of cervical length at thresholds of < 15, < 20 or < 25 mm and are based only on data reported by studies included in our systematic review of Actim Partus, PartoSure and qfFN at thresholds other than 50 ng/ml. In addition, the large variation across these studies in sensitivity and specificity may be, at least in part, owing to the different clinical personnel conducting the cervical length measurements.

### Comparison of cervical length and index tests

Six studies<sup>42–44,47,50–52</sup> assessed both cervical length measurement and Actim Partus in the same population. Three studies assessed both cervical length measurement and PartoSure<sup>44,59–61</sup> and two<sup>42,43,62</sup> assessed both cervical length and qfFN. Note that the study by Hadzi-Lega *et al.*<sup>44</sup> and the APOSTEL-1 study<sup>42,43</sup> both assess two index tests.

Against the 7-day reference standard, sensitivity was higher for cervical length measurement than Actim Partus in three studies (see Table 62).<sup>44,50,52</sup> The cervical length threshold for a positive test result was < 25 mm in Hadzi-Lega *et al.*<sup>44</sup> and Goyal *et al.*,<sup>52</sup> and < 20 mm in Danti *et al.*<sup>50</sup> In one study,<sup>47</sup> sensitivity (against the 7-day reference standard) did not differ between Actim Partus and cervical length measurement with a threshold of < 25 mm. In the remaining two studies,<sup>42,43,51</sup> sensitivity was higher for Actim Partus than for cervical length measurement with a threshold of < 15 mm. Specificity (against the 7-day reference standard) was higher for Actim Partus than for cervical length in all studies except for Danti *et al.*<sup>50</sup> and Eroglu *et al.*,<sup>51</sup> in which the specificity was higher for cervical length measurement (see Table 62).

When comparing the 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 measurement at a threshold of < 15 mm in Bolotskikh *et al.*<sup>59</sup> [100% (95% CI 73.5% to 100.0%) vs. 33.3% (95% CI 9.9% to 65.1%)] and lower for PartoSure than for cervical length at a threshold of < 25 mm in Hadzi-Lega *et al.*<sup>44</sup> [83.3% (95% CI 35.9% to 99.6%) vs. 100% (95% CI 54.1% to 100%)], although CIs overlap in Hadzi-Lega *et al.*<sup>44</sup> Conversely, specificity was lower for PartoSure than for cervical length measurement at a threshold of < 15 mm in Bolotskikh *et al.*<sup>59</sup> [95.4% (95% CI 88.6% to 98.7%) vs. 98.9% (95% CI 93.8% to 100.0%)] and higher for PartoSure than for cervical length measurement at a threshold of < 25 mm in Hadzi-Lega *et al.*<sup>44</sup> [90.2% (95% CI 78.6% to 96.7%) vs. 70.6% (95% CI 56.2% to 82.5%)], albeit with overlapping CIs. In the third study,<sup>60,61</sup> both sensitivity and specificity were higher for PartoSure than for cervical length at a threshold of < 25 mm [sensitivity 80.0% (95% CI 63.1% to 91.6%) vs. 57.1% (95% CI 39.4% to 73.7%) and specificity 94.6% (95% CI 90.1% to 97.5%) vs. 72.6% (95% CI 65.2% to 79.2%)], although, again, CIs (for sensitivity) overlap.

In comparison with qfFN, in APOSTEL-1,<sup>42,43</sup> cervical length at a threshold of < 15 mm was most closely matched to qfFN with the threshold of 200 ng/ml [sensitivity against the 7-day reference standard: 72.5% (95% CI 60.4% to 82.5%) vs. 71.0% (95% CI 58.8% to 81.3%) and specificity against the 7-day reference standard: 83.3% (95% CI 78.4% to 87.4%) vs. 83.6% (95% CI 78.8% to 87.8%)]. However, in Bruijn,<sup>62</sup> the sensitivity and specificity (against the 7-day reference standard) of cervical length measurement at a threshold of < 15 mm fell between sensitivities and specificities produced for qfFN at the 200-ng/ml and 500-ng/ml thresholds (see *Table 62*). This was particularly because sensitivity of cervical length at a threshold of < 15 mm was lower in Bruijn than in APOSTEL-1 [51.3% (95% CI 34.8% to 67.6%) vs. 72.5% (95% CI 60.4% to 82.5%)], although the 95% CIs do overlap.<sup>42,43,62</sup>

### Data from systematic reviews of cervical length measurement

In the Boots *et al.*<sup>129</sup> review, cervical length measurement at a cut-off point of 15 mm was assessed in 24 studies (against a reference standard of delivery within 7 days), with pooled sensitivity reported as 74% (95% CI 58% to 85%) and pooled specificity as 89% (95% CI 85% to 92%). Recent NICE guidance<sup>24</sup> shows how the variability across studies is great: for cervical length measurement at a cut-off point of < 15 mm, 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% to 39.7%) to 97.7% (95% CI 86.9% to 99.9%) and specificity from 83.0% (95% CI 70.0% to 93.0%) to 96.5% (95% CI 95.4% to 97.7%). These systematic review data are similar to the data for cervical length measurement from the current overview (see *Test accuracy of transvaginal cervical length measurement* and *Table 62*) in which, at a threshold of < 15 mm, sensitivity (against the 7-day reference standard) showed great variability across studies, ranging from 33.3% (95% CI 9.9% to 65.1%) to 72.5% (95% CI 60.4% to 82.5%), and specificity was more similar across studies, ranging from 81.8% (95% CI 77.7% to 85.4%) to 98.9% (95% CI 93.8% to 100.0%).

In the recent NICE guidance,<sup>24</sup> at a cut-off point of < 25 mm (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% to 64.7%) to 83.3% (95% CI 43.7% to 97.0%) and specificity ranged from 71.7% (95% CI 66.4% to 73.8%) to 96.9% (95% CI 91.6% to 99.5%). Again, these sensitivity data are similar to those in the current overview in which, at a threshold of < 25 mm, sensitivity ranged from 57.1% (95% CI 39.4% to 73.7%) to 100.0% (95% CI 54.1% to 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% to 56.6%) to 72.6% (95% CI 65.2% to 79.2%)] than in the recent NICE guidance.<sup>24</sup> The recent NICE guidance<sup>24</sup> also included test accuracy data for cervical length at a threshold of < 30 mm (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% to 97.2%) to 94.0% (95% CI 79.0% to 99.0%) and specificity from 42.0% (95% CI 37.0% to 47.0%) to 55.6% (95% CI 53.0% to 56.8%). The current overview does not provide test accuracy data for cervical length at the < 30-mm threshold.

### Summary

Overall summary tables for the DTA review, including cervical length data, are presented in *Tables 37* and *38*.



**TABLE 64** Summary of evidence and relative accuracy against the 7-day reference standard

Test	Actim Partus	PartoSure	qfFN at		
			10 ng/ml	200 ng/ml	500 ng/ml
<b>Index tests</b>					
Actim Partus					
PartoSure	No difference (Hadzi-Lega <i>et al.</i> <sup>44</sup> )				
qfFN at					
10 ng/ml	Sensitivity of fFN superior, specificity of Actim Partus superior (APOSTEL-1 <sup>42,43</sup> )	Indirect evidence only			
200 ng/ml	No difference (APOSTEL-1 <sup>42,43</sup> )		Indirect evidence only		
500 ng/ml	Sensitivity of Actim Partus superior, specificity of fFN superior (APOSTEL-1 <sup>42,43</sup> )	Indirect evidence only			
<b>fFN at 50 ng/ml</b>					
qfFN at 50 ng/ml	Specificity of Actim Partus superior, no difference in sensitivity (APOSTEL-1 <sup>42,43</sup> )	Indirect evidence only	Sensitivity of fFN at 10 ng/ml superior; specificity of fFN at 50 ng/ml superior (APOSTEL-1 <sup>42,43</sup> and Bruijn <sup>62</sup> )	Sensitivity of fFN at 50 ng/ml superior; specificity of fFN at 200 ng/ml superior (APOSTEL-1 <sup>42,43</sup> and Bruijn <sup>62</sup> )	Sensitivity of fFN at 50 ng/ml superior; specificity of fFN at 500 ng/ml superior (APOSTEL-1 <sup>42,43</sup> and Bruijn <sup>62</sup> )
QuikCheck	Sensitivity of Actim Partus superior and specificity of fFN superior (Tripathi <i>et al.</i> <sup>57</sup> ). However, Eroglu <i>et al.</i> <sup>51</sup> showed no difference between tests	Specificity of PartoSure superior, no difference in sensitivity (Nikolova <i>et al.</i> <sup>60,61</sup> ; note missing participants)	Indirect evidence only	Indirect evidence only	Indirect evidence only
ELISA	No difference (Riboni <i>et al.</i> <sup>54</sup> )	No evidence	Indirect evidence only	Indirect evidence only	Indirect evidence only

continued

**TABLE 64** Summary of evidence and relative accuracy against the 7-day reference standard (*continued*)

Test	Actim Partus	PartoSure	qfFN at		
			10 ng/ml	200 ng/ml	500 ng/ml
<b>CL<sup>a</sup></b>					
< 15 mm	No difference (APOSTEL-1 <sup>42,43</sup> )	No evidence	Sensitivity of fFN superior; specificity of CL superior (APOSTEL-1 <sup>42,43</sup> and Bruijn <sup>62</sup> )	No difference (APOSTEL-1 <sup>42,43</sup> and Bruijn <sup>62</sup> )	Sensitivity of CL superior or no difference; specificity of fFN superior (APOSTEL-1 <sup>42,43</sup> and Bruijn <sup>62</sup> )
< 20 mm	No difference (Danti <i>et al.</i> <sup>50</sup> and Eroglu <i>et al.</i> <sup>51</sup> ) <sup>b</sup>	No evidence	Indirect evidence only	Indirect evidence only	Indirect evidence only
< 25 mm	No difference (Azlin <i>et al.</i> , <sup>47</sup> Goyal <i>et al.</i> <sup>52</sup> and Hadzi-Lega <i>et al.</i> <sup>44</sup> )	Specificity of PartoSure superior or no difference; sensitivity no difference (Nikolova <i>et al.</i> <sup>60,61</sup> and Hadzi-Lega <i>et al.</i> <sup>44</sup> )	Indirect evidence only	Indirect evidence only	Indirect evidence only

CL, cervical length.

a Studies reporting accuracy of cervical length across multiple thresholds and data regarding the cut-off point closest to the 15-mm threshold (NICE guidelines<sup>24</sup>) are reported here.

b Danti *et al.*<sup>50</sup> a subset of recruited participants received the Actim Partus test, no difference between CL and Actim Partus was observed when compared in the population that received both tests.

TABLE 65 Summary table

Test	Source (first author)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)
<b>Test accuracy for the prediction of preterm delivery within 7 days</b>			
<i>Studies assessing more than one index test</i>			
fFN at 10 ng/ml	Bruijn <sup>42,43</sup> (APOSTEL-1) (n = 350)	95.7 (87.8 to 99.1)	42.3 (36.5 to 48.4)
fFN at 200 ng/ml	Bruijn <sup>42,43</sup> (APOSTEL-1) (n = 350)	71.0 (58.8 to 81.3)	83.6 (78.8 to 87.8)
fFN at 500 ng/ml	Bruijn <sup>42,43</sup> (APOSTEL-1) (n = 350)	42.0 (30.2 to 54.5)	95.7 (92.7 to 97.8)
Actim Partus	Bruijn <sup>42,43</sup> (APOSTEL-1) (n = 350)	78.3 (66.7 to 87.3)	89.3 (85.1 to 92.7)
PartoSure	Hadzi-Lega <sup>44</sup> (n = 57)	83.3 (35.9 to 99.6)	90.2 (78.6 to 96.7)
Actim Partus	Hadzi-Lega <sup>44</sup> (n = 57)	83.3 (35.9 to 99.6)	76.5 (62.5 to 87.2)
<i>Studies assessing a single index test</i>			
Actim Partus	Pooled (16 studies)	77 (68 to 83)	81 (76 to 85)
	Range (16 studies)	33.3 (4.3 to 77.7) to 94.7 (89.9 to 97.7)	50.0 (24.7 to 75.3) to 93.5 (82.1 to 98.6)
PartoSure	Pooled (4 studies)	83 (61 to 94)	95 (89 to 98)
	Range (4 studies)	0 (0.0 to 97.5) to 100.0 (73.5 to 100.0)	90.2 (78.6 to 96.7) to 97.5 (96.8 to 99.9)
fFN at 10 ng/ml	Range (2 studies)	93.8 (82.8 to 98.7) to 95.7 (87.8 to 99.1)	32.2 (27.7 to 37.0) to 42.3 (36.5 to 48.4)
fFN at 200 ng/ml	Range (2 studies)	70.8 (55.9 to 83.0) to 71.0 (58.8 to 81.3)	78.6 (74.3 to 82.5) to 83.6 (78.8 to 87.8)
fFN at 500 ng/ml	Range (2 studies)	29.2 (17.0 to 44.1) to 42.0 (30.2 to 54.5)	94.3 (91.6 to 96.4) to 95.7 (92.7 to 97.8)
<i>Supplementary data from included studies</i>			
fFN at 50 ng/ml	Range (8 studies)	23.8 (17.3 to 31.4) to 91.3 (82.0 to 96.7)	62.2 (57.3 to 66.9) to 99.1 (97.3 to 99.8)
CL < 15 mm	Range (3 studies)	33.3 (9.9 to 65.1) to 72.5 (60.4 to 82.5)	81.8 (77.7 to 85.4) to 98.9 (93.8 to 100.0)
CL < 20 mm	Danti <sup>50</sup> (n = 60)	75.0 (19.4 to 99.4)	71.4 (57.8 to 82.7)
CL < 25 mm	Range (4 studies)	57.1 (39.4 to 73.7) to 100.0 (54.1 to 100.0)	31.6 (12.6 to 56.6) to 72.6 (65.2 to 79.2)
<i>Data extracted from systematic reviews</i>			
fFN at 50 ng/ml	Sanchez-Ramos; <sup>128</sup> pooled (32 studies)	76.1 (69.1 to 81.9)	81.9 (78.9 to 84.5)
fFN at 50 ng/ml	Boots; <sup>129</sup> pooled (38 studies)	75 (69 to 80)	79 (76 to 83)
fFN at 50 ng/ml	NICE guidance; <sup>24</sup> range (20 studies)	56 <sup>a</sup> to 100 <sup>a</sup>	61.9 (59.6 to 62.5) to 92 <sup>a</sup>
CL < 15 mm	Boots; <sup>129</sup> pooled (24 studies)	74 (58 to 85)	89 (85 to 92)
CL < 15 mm	NICE guidance; <sup>24</sup> range (8 studies)	26.3 (11.2 to 39.7) to 97.7 (86.9 to 99.9)	83.0 (70.0 to 93.0) to 96.5 (95.4 to 97.7)
CL < 25 mm	NICE guidance; <sup>24</sup> range (5 studies)	60.0 (48.3 to 64.7) to 83.3 (43.7 to 97.0)	71.7 (66.4 to 73.8) to 96.9 (91.6 to 99.5)
CL < 30 mm	NICE guidance; <sup>24</sup> range (3 studies)	89.3 (71.8 to 97.2) to 94.0 (79.0 to 99.0)	42.0 (37.0 to 47.0) to 55.6 (53.0 to 56.8)

continued

TABLE 65 Summary table (continued)

Test	Source (first author)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)
<b>Test accuracy for the prediction of preterm delivery within 48 hours</b>			
<i>Studies assessing a single index test</i>			
Actim Partus	Pooled (6 studies)	87 (74 to 96)	73 (62 to 82)
	Range (6 studies)	65.7 (47.8 to 80.9) to 100 (47.8 to 100.0)	56.0 (34.9 to 75.6) to 82.4 (56.6 to 96.2)
PartoSure	Werlen <sup>a1</sup> (n = 41)	0.0 (0.0 to 97.5)	97.5 (86.8 to 99.9)
CL, cervical length. a 95% CI not reported.			



A decorative graphic consisting of numerous thin, parallel green lines that curve from the left side of the page towards the right, creating a sense of movement and depth.

**EME  
HS&DR  
HTA  
PGfAR  
PHR**

Part of the NIHR Journals Library  
[www.journalslibrary.nihr.ac.uk](http://www.journalslibrary.nihr.ac.uk)

*This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care*

***Published by the NIHR Journals Library***