

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

Jo Varley-Campbell,^{1*} Rubén Mújica-Mota,¹
Helen Coelho,¹ Neel Ocean,¹ Max Barnish,¹
David Packman,¹ Sophie Dodman,¹ Chris Cooper,¹
Tristan Snowsill,^{1,2} Tracey Kay,³ Neil Liversedge,³
Michelle Parr,⁴ Lisa Knight,³ Chris Hyde,¹
Andrew Shennan^{5,6} and Martin Hoyle¹ on behalf of
the Peninsula Technology Assessment Group (PenTAG)

¹Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, University of Exeter, Exeter, UK

²Health Economics Group, University of Exeter Medical School, University of Exeter, Exeter, UK

³Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

⁴Central Manchester University Hospital NHS Foundation Trust, Manchester, UK

⁵Department of Women and Children's Health, King's College London, London, UK

⁶Guy's and St Thomas' Hospital, London, UK

*Corresponding author j.varley-campbell@exeter.ac.uk

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

Scientific summary

Biomarker tests to diagnose preterm labour

Health Technology Assessment 2019; Vol. 23: No. 13

DOI: 10.3310/hta23130

NIHR Journals Library www.journalslibrary.nihr.ac.uk

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 ≥ 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. 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 relative to 50 ng/ml of fFN 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.

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/).

Editorial contact: journals.library@nhr.ac.uk

The full HTA archive is freely available to view online at 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

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), produced by Prepress Projects Ltd, Perth, Scotland (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

Editorial contact: journals.library@nihr.ac.uk