Development and validation of prediction models for the QUiPP App v.2: a tool for predicting preterm birth in women with symptoms of threatened preterm labor

Jenny Carter PhD[1]*, Paul T. Seed[1], Helena A. Watson[1], Anna L David[2,3], Jane Sandall PhD[1], Andrew H. Shennan, MD[1]**, Rachel M. Tribe PhD[1]**

[1]Dept. of Women and Children's Health, School of Life Course Sciences, King's College, London, London, UK

[2] Institute for Women's Health, University College London

[3] National Institutes for Health Research University College London Hospitals Biomedical Research Centre, 149 Tottenham Court Road, London W1T 7DN

*corresponding author

Department of Women and Children's Health, School of Life Course Sciences, King's College London, 10th Floor, North Wing, St Thomas' Hospital Campus, Westminster Bridge Road, London, SE1 7EH, 020 7188 3634.

** joint last authors

Short title (20 letters): QUiPP App v.2 TPTL prediction model development.

Keywords: Preterm, prediction, risk assessment, mHealth, eHealth, mobile apps

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/uog.20422

Contribution

What does this work add to what is already known?

The QUiPP v.2 app is a reliable risk assessment tool that combines risk factors with fetal fibronectin and cervical length and calculates a simple % risk of spontaneous preterm birth.

What are the clinical implications of this work?

Use of QUiPP should increase confidence in clinical decisions, improve targeting and timing of interventions to reduce preterm birth and its associated morbidities, and limit unnecessary intervention and women's anxiety.

ABSTRACT

Objective

To develop enhanced prediction models to update the QUiPP app, a tool for predicting spontaneous preterm birth in women with symptoms of threatened preterm labour (TPTL), incorporating risk factors, transvaginal ultrasound assessment of cervical length (CL, mm) and cervicovaginal fluid quantitative fetal fibronectin test results (qfFN).

Methods

Participants were pregnant women between 23^{+0} and 34^{+6} weeks' gestation with symptoms of TPTL, recruited as part of four prospective cohort studies carried out at 16 UK hospitals between October 2010 and October 2017. The training set comprised all women where outcomes were known at May 2017 (n=1032). The validation set comprised women where outcomes were gathered between June 2017 and March 2018 (n=506). Parametric survival models were developed for three combinations of predictors: risk factors plus qfFN test, risk factors plus CL only, and risk factors plus both tests. The best models were selected using the Akaike and Bayesian information criteria. The estimated probability of delivery before 30, 34 or 37 weeks' gestation and within 1 or 2 weeks of testing was calculated and Receiver Operating Characteristic (ROC) curves were created to demonstrate the diagnostic ability of the prediction models.

Results

Predictive statistics were similar in training and validation sets. Areas under the ROC curves (validation set) demonstrated good prediction at all time points, particularly in the combination of risk factors plus qfFN model: 0.96 (<30 weeks); 0.85 (<34 weeks); 0.77 (<37 weeks); 0.91 (<1 week) and 0.92 (<2 weeks).

Conclusions

Validation of these prediction models suggests the QUiPP v.2 app can reliably calculate risk of preterm delivery in women with TPTL. Use of the QUiPP app in practice could lead to better targeting of intervention, while providing reassurance and avoiding unnecessary intervention in women at low risk.

INTRODUCTION

Although advances in care have led to more, and earlier, babies surviving, preterm birth remains hard to predict, even in women with symptoms of threatened preterm labour (TPTL). Symptoms of TPTL are not accurate predictors of preterm birth[1,2] but because the consequences of not treating women in "true" preterm labour could be devastating, many receive unnecessary interventions. Overtreatment results in avoidable exposure to the danger of adverse effects, particularly with repeated doses of steroids for fetal lung maturation, as this has been associated with reduced birth weight[3-5]. It also results in significant unnecessary healthcare expenditure[6-8].

In the absence of definitive diagnostic tests for early labour, management decisions are based on assessment of risk. Risk assessment in TPTL is difficult, however, due to its multifactorial nature. Risk factors, such as previous spontaneous preterm birth (sPTB) and cervical surgery, along with gestation and nature of symptoms, need to be considered. Additional tests, e.g. transvaginal ultrasound cervical length measurement (TVS CL) and fetal fibronectin (fFN) can aid clinical decision making[9-13]. The potential of fFN as a predictive marker for sPTB has been established for several years[14-19]. In the UK, test results were, until relatively recently, presented as dichotomous (i.e. positive or negative), based on a threshold of 50 ng/ml. Newer fFN analysers provide results as concentrations in ng/ml, and it has been suggested that using alternative thresholds, i.e. <10 ng/ml and >200 ng/ml, rather than 50 ng/ml, may improve positive prediction[14,20].

We developed the QUiPP mobile phone application (www.quipp.org) for clinical decision support in women at increased risk of sPTB. This easy-to-use risk assessment tool combines risk factors and test results, and calculates a simple % risk for individual women. Algorithms used in generating the risk prediction scores for the first version of QUiPP have been previously reported[21,22]. This paper reports further development and validation of algorithms for calculating risk in women with symptoms incorporated in the second version of the app. These new algorithms improve the utility and flexibility of QUiPP through the introduction of additional risk factors and calculation of risk using either quantitative fFN (ng/ml), TVS CL (mm) or both tests combined.

METHODS

Upgrade of the QUiPP app was originally planned to include all PETRA study recruits, an observational study designed to collect data for this purpose (REC reference 14/LO/1988). However, before the participant recruitment phase was complete, a funding application to evaluate the QUiPP app's usability, acceptability and effect on management, in a clinical trial, was successful (EQUIPTT, REC Ref. 17/LO/1802; ISRCTN trial registry number ISRCTN17846337). A decision was taken to update the predictive algorithms prior to completion of PETRA using all outcome data already gathered (at the end of May 2017) along with relevant participant data, i.e. matched eligibility criteria and symptoms of TPTL, from earlier prospective cohort studies [EQUIPP (REC Ref. 10/H0806/68), POPPY (REC Ref. 09/H0802/97) and INSIGHT (REC Ref. 13/LO/0393)]. These studies all investigated preterm birth prediction and, like PETRA, utilised our Preterm Birth Studies database (www.medscinet.net/ptbstudies). A summary of study characteristics and number of participants from each study included in the training set is shown in Figure 1.

Participants were pregnant women between 23^{+0} and 34^{+6} weeks gestation with symptoms of threatened preterm labour (e.g. abdominal pain or tightenings). Women were excluded if diagnosed with established labour, ruptured membranes or antepartum haemorrhage. Fetal fibronectin test results and cervical length measurements were known to attending clinicians and management, e.g. hospital admission, administration of steroids and tocolytics, was according to local protocols. All data was gathered between October 2010 and October 2017 from 16 UK hospitals.

The training set for this analysis (n=1032) included data from 382 EQUIPP study participants (37%) that had been used for development of the first QUiPP symptomatic algorithms[21], although this was limited to data relating to women with only fFN test results and singleton pregnancies. As there was no statistical reason to exclude these data from the current analysis, they were incorporated in order to increase the predictive ability of the new models. Additional data on cervical length and twin pregnancies, collected since creation of the first prediction models, increases the flexibility of the QUiPP app in practice. The new models were tested by

calibration before being applied in the second version of the QUiPP app. CE Marking as a Class 1 Medical Device was granted before general release in September 2017. (MHRA Ref. no. for Medical Device/standalone software Z301 registration is A015030). Formal validation of the prediction models was carried out after completion of the PETRA study (April 2018, n=506).

Power calculation

A power calculation was performed prior to commencement of the PETRA study. We anticipated that clinicians would be willing to view women in the lower risk group as closer to the normal (i.e. standard risk) if the true rate of PTL in this group could be demonstrated (with 95% confidence) to be lower than the expected rate for women with TPTL symptoms (i.e. lower than 10% with a best estimate of 6.7%), and concluded that full data on 550 standard risk women and 61 high risk women (total 611 women) in the proposed validation would be sufficient to achieve 80% power in the PETRA study. Allowing for 95% compliance & completion, a recruitment target of 643 women was considered adequate to validate the predictive value of each test (qfFN and CL) with an additional 300 to be used as a training set. The final numbers used in the development and validation of the new algorithms provide sufficient power to achieve our objectives.

Predictive model generation

In total, six prediction algorithms were needed for the development of the new version (v.2) of the QUiPP app: three for symptomatic women, three for asymptomatic women. The algorithm is selected according to whether the woman is asymptomatic high risk (e.g. with a previous history of preterm birth, preterm prelabour ruptured membranes, late miscarriage or cervical surgery) or symptomatic of TPTL (any risk status) and whether her risk assessment includes qfFN concentration alone, CL measurement alone, or both test results. Data were therefore split and tested in six groups: asymptomatic high risk with: i) qfFN test; ii) CL measurement; iii) both test results; and symptomatic (any risk status) with iv) qfFN test; v) CL measurement; vi) both test results. In this paper, we report findings from the validation of the three algorithms appropriate for the cohorts symptomatic of TPTL. Development of the prediction models for asymptomatic high risk women are reported in our complementary paper (*submitted to this journal on ##/04/2019*).

Statistical analysis was carried out using Stata SE software (version 14.2; StataCorp, College Station, Texas, USA). Exclusions were made for: incomplete data; invalid visits (out of gestation range, inappropriate symptoms, invalid or missing test results, sexual intercourse within 24 hours) and major fetal abnormality (Figure 2). Women with twin pregnancies were included, using the first twin gestation at delivery, but triplets and higher order multiples were excluded due to inadequate numbers. Women whose labour was induced or who had caesarean section following preterm prelabour rupture of membranes (PPROM) were regarded as having had spontaneous preterm birth. In multiple pregnancies, outcomes for the first baby were used in the analysis.

Women who had received interventions that are intended to reduce the risk of sPTB (i.e. tocolysis, progesterone, cerclage and Arabin pessary) were not excluded in the models. This is justified because, as this was not a randomised trial, any estimated treatment differences were likely to have been misleading, and could even have been in the wrong direction when compared to the true treatment effect. A treatment paradox operates: typically the more severe the illness, the more likely that a clinical decision will be made to use a particular treatment. A decision to treat to prevent early delivery is therefore a marker for greater risk of early delivery, even if the treatment itself tends to lengthen gestation. Looking solely at untreated women would mean selecting the women perceived as at lowest risk, and so underestimating the risk of prematurity. Including all women, irrespective of treatment, while not ideal, gives us the best estimate of risk in the clinical setting where the data was gathered. Additionally, if we include the changes associated with treatment in the app, there would be a real risk of clinicians using it to decide not to treat, even when treatment would be beneficial.

Cox's proportional hazards regression was used to determine which predictive risk factors to use in the model. Factors tested included demographic characteristics, i.e. age, Body Mass Index (BMI, kg/m^2), ethnicity, deprivation score and smoking), clinical risk factors (i.e. previous history of preterm birth or PPROM, late miscarriage, cervical surgery, twin pregnancy) and test results (qfFN and TVS CL). Simple regression methods were not sophisticated enough for creation of the QUiPP app prediction models because time to delivery after testing has to be very precise, with very smooth survival curves, and therefore parametric survival analysis was

used. This process involved testing the data using several different parametric survival analysis functions, namely exponential, gamma, Gompertz, log-logistic, log-normal and Weibull.

Where women presented more than once in a pregnancy for TPTL assessment, later results were introduced as time-updated covariates, i.e. if delivery has not occurred before the next visit, prediction was recalculated with the next visit gestation. In survival analysis, data are "censored" if the outcome of interest has not occurred during the follow up period[23]. In this study, the data were censored, if spontaneous preterm birth had not occurred by 37 weeks' gestation. Iatrogenic preterm birth was treated as a non-event so data relating to women with this pregnancy outcome were censored at term. Checks were undertaken to determine whether the data needed to be transformed before analysis using fractional polynomials. The entire procedure was repeated for each of the three combinations of predictors (i.e. risk factors plus qfFN test, risk factors plus CL only, and risk factors plus both tests), and different models were produced in each case. The best models were then determined by reference to Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC), where the lowest values are considered to have the best fit to the data[24]. This is a method developed for comparing non-nested regression models where significance tests are not available.

None of the predetermined demographic factors, i.e. age, BMI, ethnicity, deprivation score and smoking, affected prediction of sPTB in the models. This was not because they have no value as predictors in themselves, but that the other predictors (i.e. major risk factors, qfFN and CL) were much stronger, so adding them did not affect the overall score. In testing the model with the cohort of women who had both fFN and CL test results, multivariate regression showed that only previous cervical surgery provided additional predictive power to fFN and CL test results in women with symptoms of TPTL. However, the composite of risk factors used in the asymptomatic prediction algorithm for the QUiPP v.2 app, which was being developed in tandem, (i.e. multiple pregnancy, history of sPTB or PPROM, late miscarriage or cervical surgery) was tested to establish whether it affected the prediction in the symptomatic women. There was little difference, so a decision was made to use this composite of risk factors for consistency.

The prediction models were then tested by simple calibration. This meant comparing individual tests of clinically significant groups to confirm the actual event rates were consistent with the predicted probability of the event. A 5% prediction rate for sPTB within 7 days of testing was

used as the threshold because this was the lowest value of a range of 5-15% that our TPTL Delphi consensus survey suggested should be recommended for intervention[25]. The calibration tests provided reassurance that the models were acceptable to proceed with development of the QUiPP app before formal validation was undertaken.

Predictive model validation

Validation was carried out on a later subset of women from the PETRA study (n=506). Predictive statistics, including sensitivity, specificity, balanced accuracy [(sens.+spec.)/2], likelihood ratios, positive (PPV) and negative predictive values (NPV) and separation probabilities (PPV+NPV-100%) were calculated using a % risk of ≥5% as an indication of a positive test. This threshold was chosen, as it was for the calibration exercise, with reference to our TPTL Delphi consensus survey[25]. Results are tabulated with statistics for both the training and validation sets, by test group (risk factors plus either qfFN, CL or both tests) for prediction of spontaneous preterm birth at less than 30, 34 and 37 weeks' gestation, and within 1 and 2 weeks post-test. These time points were chosen because: i) the gestations at delivery are clinically important indicators for likely neonatal morbidity and ii) they are useful in guiding appropriate management, such as the timing of steroids. Receiver operating characteristic (ROC) curves were drawn and areas under the curve (AUC) were calculated.

RESULTS

Participant characteristics

As explained above, three algorithms were developed so that the QUiPP app could be used in different symptomatic TPTL scenarios, i.e. when a women has: i) fFN testing alone, ii) CL measurement alone, or iii) both tests. After exclusions, as shown in Figure 2, the training dataset comprised 1173 observations from 1032 women with fFN test results, and 229 observations from 204 women with both qfFN and CL. The validation set comprised observations involving 576 qfFN tests, (506 women), 155 CL measurements (132 women) and 143 observations that include both qfFN and CL measurements (128 women). The training set included 41 sets of twins, while the validation set included 32 sets. Participant demographic characteristics and risk status by training and validation sets are shown in Tables 1 and 2.

Where intervention status was known, 30.3% (310/1024) women received steroids for fetal lung maturation, 8.2% (115/1405) received tocolysis. Most women (92.4%, 1292/1399) received no prophylactic intervention for preterm birth risk (e.g. cerclage or progesterone).

The prevalence of outcomes of sPTB at 30, 34 and 37 weeks' gestation, and at or less than 1 and 2 weeks of test for each of the test groups and between training and validation sets, were similar, as shown in Table 3.

Predictive statistics

The prediction models created generated formulae that provide individual risk scores dependent on risk factors and test results. Tables 4, 5 and 6 show predictive statistics when the algorithms are tested on both the training and validation sets, by test group for prediction of sPTB at less than 30, 34 and 37 weeks' gestation, and within 1 and 2 weeks post-test.

Tables 4, 5 and 6 show a reasonable similarity between the training and validation sets at most outcome time points and for each combination of predictors. In the qfFN group (the largest group), the ability of the algorithm to predict sPTB at less than 30 weeks' gestation had the highest balanced accuracy with, in the validation set, a sensitivity of 90.0%, specificity

of 90.8%, a positive likelihood ratio (LR+) of 9.83, a negative likelihood ratio (LR-) of 0.11, positive predictive value (PPV) of 27.3% and a negative predictive value (NPV) of 99.6%. Although NPV will always be high where prevalence is low, when NPV is greater than the overall proportion of women unaffected, as it is with all combinations of predictors for sPTB <30 weeks' (risk factors plus: qfFN 99.6% vs 96.3%; CL 98.1% vs 90.2%; both tests 100% vs 90.4%), these findings demonstrate the usefulness of QUiPP as a predictive test.

While the balanced accuracy statistics noted in Tables 4, 5 and 6 reflect the balance of sensitivity and specificity using the ≥5% risk cut off, the ROC curves shown in Figures 3, 4 and 5 indicate overall test performance, using validation set only, at all percentage risks (i.e. without using ≥5% as a cut off for positive test). The Area under the ROC curve (AUC) is a measure of how well the parameter (% risk) can distinguish between two groups, in this case, women with or without a pregnancy outcome of sPTB.

For the qfFN group, the AUC for predicting sPTB at less than 30 weeks' indicates good prediction, at 0.96, with similarly large AUCs for predicting sPTB at less than 1 week and 2 weeks post test. The risk prediction algorithm using cervical length appears to perform best at prediction of sPTB < 30 weeks, but this is inferior to the qfFN test. When both test results are combined, the prediction improves, but is inferior to qfFN alone at all time points.

In order to directly compare predictive ability of the different combinations of predictors we compared AUCs in the validation set of women who had had both tests (Figure 6 and Table 7).

Although the addition of CL to qfFN appears to be useful, the comparisons as shown in Table 7, indicate there is no difference between the individual tests or combination of both tests for predicting sPTB at 30, 34 or 37 weeks. However, at 1 and 2 weeks post test, qfFN alone appears to be a better predictor than CL alone, but it was no better than combined qfFN and CL. CL alone, however, has reduced ability to predict sPTB, with AUCs of 0.6975 and 0.7306, respectively. The number of women in this cohort having both tests was small (particularly so for prediction of sPTB at less than 30 weeks) so these results must be interpreted with caution.

As model development included data from women with interventions for reducing likelihood of sPTB (i.e. tocolysis, progesterone, cerclage and Arabin pessary), we wanted to confirm risk was not underestimated. Consequently, we compared the models between women who had these interventions with those who had not. We found the app either performed similarly in women with intervention (compared to those without) or showed significantly poorer agreement (i.e. smaller AUC) or higher risk (OR > 1 by logistic regression).

DISCUSSION

In this study, we have demonstrated an improved ability of QUiPP algorithms to predict sPTB compared to those algorithms created for the first version of the app[21]. The previous version included data from 382 women (190 training set, 192 validation set) with only the combination of risk factors and qfFN test results, as TVS CL data was unavailable. Predictive statistics demonstrated the ability of the model to predict sPTB at <30, <34 and <37 weeks, and within 2 and 4 weeks, using a threshold of >10% as a positive result. Using new algorithms created for QUiPP v.2, prediction was investigated for sPTB at <30, <34 and <37 weeks, and within 1 and 2 weeks, using a threshold of ≥5%, with a substantially larger cohort (1032 training set and 506 validation set). Although comparison can only be made with the qfFN group predictive statistics, our findings demonstrate a significant increase in sensitivity, the test's ability to correctly predict sPTB, at all outcome time points. Positive predictive values (PPV), i.e. probability that a woman with a positive test (in this case, a ≥ 5 or 10% risk of sPTB) will have sPTB, are lower in the later cohort while the negative predictive values (NPV), i.e. probability that a woman with a negative test (% risk < 5% or 10%) will not have sPTB, is similar to the Kuhrt[21] cohort. Unlike sensitivity and specificity, PPV and NPV are dependent on prevalence, and NPV will always be high where the prevalence is low. In Kuhrt's study, prevalence of sPTB was higher at all time points, so it is not surprising the PPV is higher than in the validation set.

The AUCs demonstrating the predictive ability of QUiPP v.2 algorithms can also be compared with our previous findings[21]. In this earlier study, we found AUCs of 0.88, 0.83, 0.77, 0.77 and 0.78, for prediction of sPTB at <30, <34, <37 week's gestation and within two and four weeks of testing, respectively. This represented an overall improvement compared to an earlier systematic review of fFN for predicting sPTB [15]. Honest *et al*.'s review included data from forty studies and 26,876 women, in which ROC curves ranging from 0.71 to 0.77 demonstrated the ability of fFN to predict sPTB at < 34 and <37 weeks' respectively. In the QUiPP v.2 validation set, AUCs were higher than previously reported[15,21] in all but the <37 weeks' time point: 0.96 (<30 weeks); 0.85 (<34 weeks); 0.77 (<37 weeks); 0.91 (<1 week) and 0.92 (<2 weeks).

Comparison of predictive statistics with our earlier work[21] demonstrate improved prediction but are based on algorithms developed using risk factors and qfFN only. QUiPP v.2 algorithms were created and validated for predicting sPTB using risk factors in combination with either qfFN, TVS CL or both tests. The ability of QUiPP v.2 to predict sPTB using risk factors and either, or both tests, increases its utility and flexibility as it can be used where fFN testing is unavailable, and TVS CL is increasingly common as training becomes more widespread.

When we compared the models between women who had interventions to reduce risk of sPTB with those who had not we found little difference, or reduced AUCs in the higher risk group receiving interventions. This reduction in AUC is typically found when comparing homogenous subgroups and does indicate poor model performance.

Limitations

Similar to other studies of TPTL, prevalence of preterm birth was low. Only 17 women (1.6%) of the total PETRA cohort (n=1037) delivered < 30 weeks' gestation. One reason for the low prevalence was because the prospective design meant women had to be recruited before the outcome (gestation at delivery) was known. Many women whose TPTL symptoms progressed quickly into established labour could have been missed because research staff were unable to approach them before they delivered. Despite the low prevalence, however, the overall cohort size is larger than previously reported, so the number of events is greater, which allows for increased confidence in the findings.

Implications for practice

The UK's National Institute for Health and Care Excellence (NICE) Preterm Birth guideline[26] recommends that, in women over 30 weeks' gestation, TVS CL should be offered first, followed by fFN testing, only if TVS CL is unavailable. Combining both tests is not recommended. While some investigators have found added value in combining tests[27-31], others have not[32,33]. In this project, the effect of combining CL with qfFN on predictive ability was also examined. Results indicated prediction was not improved and, indeed, that TVS CL alone was inferior in predicting sPTB within 1 or 2 weeks (AUC 0.698 vs 0.875, sPTB < 1 week,

p=0.01; AUC 0.731 vs 0.889, sPTB < 2 weeks, p=0.02). This suggests fFN has superior predictive ability, and based on these findings, fFN should be recommended as first choice of test in TPTL over TVS CL.

For women with suspected PTL under 30 weeks' gestation, NICE recommends a "treat all" strategy, without reference to either fFN or CL tests. We modelled the effect of this strategy on a cohort of 188 symptomatic women < 30 weeks', using the QUiPP app, and found that 89% (n=169) of hospital admissions could have been safely avoided if a threshold of 5% risk of delivery within the seven days had been used to guide clinical practice[34].

In conclusion, QUiPP v.2 is a reliable, simple-to-use tool, which combines risk factors and test results into one simple % risk score. Its use could increase confidence in management decisions and lead to improved targeting and timing of interventions for reducing sPTB and its associated morbidities, while limiting unnecessary intervention and women's anxiety. The ability of the new algorithms to predict sPTB < 30 weeks are particularly important and should inform revision of the current NICE "treat all" < 30 weeks strategy. Results of the EQUIPTT trial[35] may also provide evidence for review of the NICE recommendations.

ACKNOWLEDGEMENTS

The authors wish to thank all the women who took part in this study and Tommy's charity which supports all the research in the Department of Women and Children's Health at St Thomas' Hospital. This is a summary of independent research funded by the National Institute for Health Research (NIHR)'s NIHR/HEE CAT Clinical Doctoral Research Fellowship Programme (Ref.CDRF-2013-04-026). Paul Seed is partly funded by Tommy's (Registered Charity No. 1060508) and by NIHR Collaboration for Leadership in Applied Health Research and Care, South London. Jane Sandall is a National Institute for Health Research (NIHR) Senior Investigator and also supported by the NIHR Collaboration for Leadership in Applied Health Research and Care South London at King's College Hospital NHS Foundation Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. This research is supported by the NIHR BRC at Guy's & St Thomas' NHS Foundation Trust and King's College, London, and at University College London Hospitals NHS Foundation Trust and University College London.

REFERENCES

1. Copper RL, Goldenberg RL, Davis RO, Cutter GR, DuBard MB, Corliss DK, Andrews JB. Warning symptoms, uterine contractions, and cervical examination findings in women at risk of preterm delivery. *Am J Obstet Gynecol.* 1990*;* 162(**3**): 748-754.
2. Iams JD, Newman RB, Thom EA, Goldenberg RL, Mueller-Heubach E, Moawad A, Sibai BM, Caritis SN, Miodovnik M, Paul RH. Frequency of uterine contractions and the risk of spontaneous preterm delivery. *N Engl J Med*. 2002; 346(**4**): 250-255.
3. Asztalos E, Willan A, Murphy K, Matthews S, Ohlsson A, Saigal S, Armson A, Kelly E, Delisle M, Gafni A. Association between gestational age at birth, antenatal corticosteroids, and outcomes at 5 years: multiple courses of antenatal corticosteroids for preterm birth study at 5 years of age (MACS-5). *BMC Pregnancy Childbirth.* 2014*;* 14(**1**): 272.
4. Murphy KE, Hannah ME, Willan AR, Hewson SA, Ohlsson A, Kelly EN, Matthews SG, Saigal S, Asztalos E, Ross S. Multiple courses of antenatal corticosteroids for preterm birth (MACS): a randomised controlled trial. *Lancet. 2008;* 372(**9656**): 2143-2151.
5. Norberg H, Stålnacke J, Nordenström A, Norman M. Repeat Antenatal Steroid Exposure and Later Blood Pressure, Arterial Stiffness, and Metabolic Profile. *J Pediatr.* 2013*;* 163(**3**): 711-716.
6. Lucovnik M, Chambliss LR, Garfield RE. 2013, "Costs of unnecessary admissions and treatments for “threatened preterm labor.” *Am J Obstet Gynecol. 2013;* 209(**3**): 217.e1-217.e3.
7. Mozurkewich EL, Naglie G, Krahn MD, Hayashi RH. Predicting preterm birth: a cost-effectiveness analysis. *Am J Obstet Gynecol.* 2000; 182(**6**): 1589-1598.
8. van Baaren G, Vis JY, Grobman WA, Bossuyt PM, Opmeer BC, Mol BW. Cost-effectiveness analysis of cervical length measurement and fibronectin testing in women with threatened preterm labor. *Am J Obstet Gynecol.* 2013; 209(**5**): 436.e1-436.e8.
9. Fuchs IB, Henrich W, Osthues K, Dudenhausen JW. Sonographic cervical length in singleton pregnancies with intact membranes presenting with threatened preterm labor. *Ultrasound Obstet Gynecol.* 2004; 24(**5**): 554-557.
10. Iams JD, Goldenberg RL, Meis PJ, Mercer BM, Moawad A, Das A, Thom E, McNellis D, Copper RL, Johnson F. The length of the cervix and the risk of spontaneous premature delivery. *N Engl J Med.* 1996; 334(**9**): 567-573.
11. Leitich H, Brunbauer M, Kaider A, Egarter C, Husslein P. Cervical length and dilatation of the internal cervical os detected by vaginal ultrasonography as markers for preterm delivery: A systematic review. *Am J Obstet Gynecol.* 1999; 181(**6**): 1465-1472.

12. Owen J, Iams JD, National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. What we have learned about cervical ultrasound. *Semin Perinatol.* 2003; 27(**3**): 194-203.

13. Sotiriadis A, Papatheodorou S, Kavvadias A, Makrydimas G. Transvaginal cervical length measurement for prediction of preterm birth in women with threatened preterm labor: a meta-analysis. *Ultrasound Obstet Gynecol.* 2010; 35(**1**): 54-64.

14. Abbott DS, Radford SK, Seed PT, Tribe RM, Shennan AH. Evaluation of a quantitative fetal fibronectin test for spontaneous preterm birth in symptomatic women. *Am J Obstet Gynecol.* 2013; 208(**2**): 122. e1-122. e6.

15. Honest H, Bachmann LM, Coomarasamy A, Gupta JK, Kleijnen J, Khan KS. Accuracy of cervical transvaginal sonography in predicting preterm birth: a systematic review. *Ultrasound Obstet Gynecol.* 2002; 22(**3**): 305-322.

16. Leitich H, Egarter C, Kaider A, Hohlagschwandtner M, Berghammer P, Husslein P. Cervicovaginal fetal fibronectin as a marker for preterm delivery: a meta-analysis. *Am J Obstet Gynecol.* 1999; 180(**5**): 1169-1176.

17. Lockwood CJ, Senyei AE, Dische MR, Casal D, Shah KD, Thung SN, Jones L, Deligdisgh L, Garite TJ. Fetal fibronectin in cervical and vaginal secretions as a predictor of preterm delivery. *N Engl J Med.* 1991; 325(**10**):669-674.

18. Matsuura H, Takio K, Titani K, Greene T, Levery SB, Salyan ME, Hakomori S. The oncofetal structure of human fibronectin defined by monoclonal antibody FDC-6. Unique structural requirement for the antigenic specificity provided by a glycosylhexapeptide. *J Biol Chem.* 1988; 263(**7**):3314-3322.

19. Peaceman AM, Andrews WW, Thorp JM, Cliver SP, Lukes A, Iams JD, Coultrip L, Eriksen N, Holbrook RH, Elliott J, Ingardia C, Pietrantoni M. Fetal fibronectin as a predictor of preterm birth in patients with symptoms: A multicenter trial. *Am J Obstet Gynecol.* 1997; 177(**1**):13-18.

20. Foster C, Shennan AH. Fetal fibronectin as a biomarker of preterm labor: a review of the literature and advances in its clinical use. *Biomark Med.* 2014; 8(**4**): 471-484.

21. Kuhrt K, Hezelgrave N, Foster C, Seed PT, Shennan AH. Development and validation of a tool incorporating quantitative fetal fibronectin to predict spontaneous preterm birth in symptomatic women. *Ultrasound Obstet Gynecol.* 2016; 47(**2**): 210-216.

22. Kuhrt K, Smout E, Hezelgrave N, Seed PT, Carter J, Shennan AH. Development and validation of a tool incorporating cervical length and quantitative fetal fibronectin to predict spontaneous preterm birth in asymptomatic high-risk women. *Ultrasound Obstet Gynecol.* 2016; 47(1): 104-109.

23. Kleinbaum DG, Klein M. *Survival analysis (Vol. 3*). New York: Springer; 2010.

24. Royston P, Sauerbrei W. *Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables.* London: John Wiley & Sons; 2008.

25. Carter J, Tribe R, Watson H, Shennan AH. Threatened preterm labour management: results of a Delphi consensus on best practice: PL 37. *Br J Obstet Gynaecol.* 2016 Apr 1; 123:100-101.

26. National Institute for Health and Care Excellence. *Preterm labour and birth*. NICE guideline (NG25). 2015.

27. Bolt LA, Chandiramani M, De Greeff A, Seed PT, Kurtzman J, Shennan AH. The value of combined cervical length measurement and fetal fibronectin testing to predict spontaneous preterm birth in asymptomatic high-risk women. *J Matern Fetal Neonatal Med.* 2011; 24(**7**): 928-932.

28. Bruijn MM, Vis J, Wilms FF, Oudijk M, Kwee A, Porath MM, Oei G, Scheepers H, Spaanderman ME, Bloemenkamp K. Quantitative fetal fibronectin testing in combination with cervical length measurement in the prediction of spontaneous preterm delivery in symptomatic women. *Br J Obstet Gynaecol.* 2016; 123(**12**): 1965-1971.

29. 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. *Am J Obstet Gynecol.* 2013; 208(**3**): 233.e1-233.e6.

30. Gomez R, Romero R, Medina L, Nien JK, Chaiworapongsa T, Carstens M, Gonzalez R, Espinoza J, Iams JD, Edwin S, Rojas I. Cervicovaginal fibronectin improves the prediction of preterm delivery based on sonographic cervical length in patients with preterm uterine contractions and intact membranes. *Am J Obstet Gynecol.* 2005; 192(**2**): 350-359.

31. 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. *Am J Obstet Gynecol.* 2007; 197(**4**): 426.e1-426.e7.

32. Levine LD, Downes KL, Romero JA, Pappas H, Elovitz MA. Quantitative fetal fibronectin and cervical length in symptomatic women: results from a prospective blinded cohort study. *J Matern Fetal Neonatal Med.* 2018 May 15: 1-9.

33. Tsoi E, Akmal S, Geerts L, Jeffery B, Nicolaides K. Sonographic measurement of cervical length and fetal fibronectin testing in threatened preterm labor. *Ultrasound Obstet Gynecol.* 2006; 27(**4**):368-372.

34. Watson HA, Carter J, Seed PT, Tribe RM, Shennan AH. The QUiPP App: a safe alternative to a treat-all strategy for threatened preterm labor. *Ultrasound Obstet Gynecol*. 2017; 50(**3**): 342-346.

35. Watson HA, Carlisle N, Kuhrt K, Tribe RM, Carter J, Seed P, Shennan AH. EQUIPTT: The Evaluation of the QUiPP app for Triage and Transfer protocol for a cluster randomised trial to evaluate the impact of the QUiPP app on inappropriate management for threatened preterm labour. *BMC Pregnancy Childbirth*. 2019; 19(**1**): 68.

FIGURES

Figure 1: Summary of study characteristics and number of participants from each study included in the training and validation sets.

Figure 2: Flow chart showing participants after exclusions and split between training and validation sets. TPTL = threatened preterm labour, qfFN = quantitative fetal fibronectin, CL = cervical length

Figure 3. ROC curves showing QUiPP app prediction of spontaneous preterm birth (sPTB) at less than 30, 34 and 37 weeks' gestation and within 1 and 2 weeks of testing in the group of women with fetal fibronectin (qfFN) test results in the validation set. AUC=area under the ROC curve.

Figure 4. ROC curves showing QUiPP app prediction of spontaneous preterm birth (sPTB) at less than 30, 34 and 37 weeks' gestation and within 1 and 2 weeks of testing in the group of women with cervical length (CL) in the validation set. AUC=area under the ROC curve.

Figure 5. ROC curves showing QUiPP app prediction of spontaneous preterm birth (sPTB) at less than 30, 34 and 37 weeks' gestation and within 1 and 2 weeks of testing in the group of women with both fetal fibronectin (qfFN) and cervical length (CL) in the validation set. AUC=area under the ROC curve.

Figure 6. ROC curves showing ability of QUiPP app to predict spontaneous preterm birth (sPTB) at less than 30, 34 and 37 weeks' gestation and within 1 and 2 weeks of testing in the group of women with both fetal fibronectin (fFN) and cervical length (CL) in the validation set, based on qfFN alone, CL alone, or combination of both tests. AUC=area under the ROC curve.

Table 1. Demographic characteristics of women in the training and validation sets.

		Training set†		**Validation set†**		**Both groups combined†**	
AGE	n=	1032		506		1538	
	mean *(SD)*	29.9	*(5.7)*	29.8	*(6.0)*	29.9	*(5.8)*
BMI (kg/m2)	n=	1025		506		1531	
	mean *(SD)*	26.1	*(5.9)*	26.0	*(6.1)*	26.1	*(5.9)*
IMD deprivation score	n=	947		504		1451	
	mean *(SD)*	31.3	*(13.5)*	30.2	*(15.2)*	30.1	*(14.1)*
ETHNICITY	n=	1024 *(%)*		506 *(%)*		1530 *(%)*	
	European	562 *(54.9)*		326 *(64.4)*		888 *(58.0)*	
	African or Caribbean	70 *(6.8)*		33 *(6.5)*		103 *(6.7)*	
	Asian (India/Pakistan/Bangladesh)	277 *(27.1)*		92 *(18.2)*		369 *(24.1)*	
	Other (incl. Chinese)	115 *(11.2)*		55 *(10.9)*		170 *(11.1)*	

SD=standard deviation; BMI=Body Mass Index; IMD=Index of Multiple Deprivation.

†Numbers in groups differ when data is missing.

Table 2. Major risk factors of women in training and validation sets.

	Training set n=1032	**Validation set n=506**	**Both groups combined n=1538**
	n *(%)*	n *(%)*	n *(%)*
Previous PTB < 37 weeks	158 *(15.3)*	83 *(16.4)*	241 *(15.7)*
Previous PPROM < 37 weeks	74 *(7.2)*	34 *(6.7)*	108 *(7.0)*
Previous late miscarriage	79 *(7.7)*	13 *(2.6)*	92 *(6.0)*
Cervical surgery	65 *(6.3)*	26 *(5.1)*	91 *(5.9)*
Twin pregnancy	41 *(4.0)*	33 *(6.5)*	74 *(4.8)*

PTB=preterm birth; PPROM=prelabour preterm ruptured membranes.

Table 3. Number of tests in each of the test groups with outcomes by training and validation sets and spontaneous preterm birth (sPTB) at less than 30, 34 and 37 weeks' gestation and within 1 and 2 weeks of testing.

qfFN group						
	Training set			Validation set		
sPTB	Total	n=sPTB *(%)*	*95% CI*	Total	n=sPTB *(%)*	*95% CI*
<30wk*	574	22 *(3.8)*	*2.4-5.7*	272	10 *(3.7)*	*1.8-6.7*
<34wk**	1066	60 *(5.6)*	*4.3-7.2*	520	26 *(5.0)*	*3.3-7.2*
<37wk	1173	144 *(12.3)*	*10.5-14.3*	576	68 *(11.8)*	*9.3-14.7*
<1wk	1173	15 *(1.3)*	*0.7-2.1*	576	13 *(2.3)*	*1.2-3.8*
<2wk	1173	38 *(3.2)*	*2.3-4.4*	576	18 *(3.1)*	*1.9-4.9*
CL group						
	Training set			Validation set		
sPTB	Total	n=sPTB *(%)*	*95% CI*	Total	n=sPTB *(%)*	*95% CI*
<30wk*	147	17 *(11.6)*	*6.9-17.9*	92	9 *(9.8)*	*4.6-17.8*
<34wk**	214	41 *(19.2)*	*14.1-25.1*	150	17 (*11.3)*	*6.7-17.5*
<37wk	229	69 *(30.1)*	*24.3-36.5*	155	32 *(20.6)*	*14.6-27.9*
<1wk	229	8 *(3.5)*	*1.5-6.8*	155	7 *(4.5)*	*1.8-9.1*
<2wk	229	21 *(9.2)*	*5.8-13.7*	155	8 *(5.2)*	*2.3-9.9*
Both qfFN and CL group						
	Training set			Validation set		
sPTB	Total	n=sPTB *(%)*	*95% CI*	Total	n=sPTB *(%)*	*95% CI*
<30wk*	147	17 *(11.6)*	*6.9-17.9*	83	8 *(9.6)*	*4.3-18.1*
<34wk**	214	41 *(19.2)*	*14.1-25.1*	138	16 *(11.6)*	*6.8-18.1*
<37wk	229	69 *(30.1)*	*24.3-36.5*	143	31 *(21.7)*	*15.2-29.3*
<1wk	229	8 *(3.5)*	*1.5-6.8*	143	7 *(4.9)*	*2.0-9.8*
<2wk	229	21 *(9.2)*	*5.8-13.7*	143	8 *(5.6)*	*2.4-10.7*

*some women were recruited after 30 weeks therefore not included here.

**some women were recruited after 34 weeks therefore not included here.

Table 4. Predictive statistics for spontaneous preterm birth (sPTB) at less than 30, 34 and 37 weeks' gestation and within 1 and 2 weeks of testing in the group of women with cervicovaginal fluid fetal fibronectin (qfFN) tests by training and validation sets.

qfFN group	**sPTB at less than**						**sPTB within**			
Outcome	**30 wk**	***95% CI***	**34 wk**	***95% CI***	**37 wk**	***95% CI***	**1 wk**	***95% CI***	**2 wk**	***95% CI***
Sensitivity %										
Training	81.8%	*(59.7-94.8%)*	80.0%	*(67.7-89.2%)*	82.6%	*(75.4-88.4%)*	73.3%	*(44.9-92.2%)*	81.6%	*(65.7-92.3%)*
Validation	90.0%	*(55.5-99.7%)*	84.6%	*(65.1-95.6%)*	80.9%	*(69.5-89.4%)*	53.8%	*(25.1-80.8%)*	83.3%	*(58.6-96.4%)*
Specificity %										
Training	92.9%	*(90.5-94.9%)*	74.2%	*(71.3-76.8%)*	62.7%	*(59.6-65.6%)*	94.0%	*(92.4-95.3%)*	86.6%	*(84.5-88.5%)*
Validation	90.8%	*(86.7-94.0%)*	70.9%	*(66.6-74.8%)*	56.9%	*(52.5-61.2%)*	92.0%	*(89.5-94.1%)*	84.2%	*(80.9-87.2%)*
Balanced accuracy *((Sens.+Spec)/2)*										
Training	87.38%	*(76.66-93.58%)*	77.08%	*(71.41-81.91%)*	72.66%	*(68.98-76.05)*	83.64%	*(68.78-92.23%)*	84.09%	*(76.81-89.41%)*
Validation	90.42%	*(75.96-96.57%)*	77.73%	*(69.54-84.22%)*	68.89%	*(63.28-73.99%)*	72.93%	*(56.27-84.94%)*	83.78%	*(73.07-90.77%)*
Likelihood ratio - positive										
Training	11.58	*(8.07-16.62)*	3.10	*(2.63-3.65)*	2.21	*(1.99-2.47)*	12.13	*(8.29-17.75)*	6.09	*(4.93-7.53)*
Validation	9.83	*(6.37-15.16)*	2.90	*(2.34-3.60)*	1.88	*(1.61-2.19)*	6.74	*(3.79-11.98)*	5.28	*(3.99-7.00)*
Likelihood ratio - negative										
Training	0.20	*(0.08-0.47)*	0.27	*(0.16-0.45)*	0.28	*(0.19-0.40)*	0.28	*(0.12-0.66)*	0.21	*(0.11-0.42)*
Validation	0.11	*(0.02-0.71)*	0.22	*(0.09-0.54)*	0.34	*(0.20-0.55)*	0.50	*(0.28-0.90)*	0.20	*(0.07-0.56)*
Positive Predictive Value (%)										
Training	31.6%	*(19.9-45.2%)*	15.6%	*(11.7-20.1%)*	23.7%	*(20.0-27.6%)*	13.6%	*(7.0-23.0%)*	16.9%	*(11.8-23.2%)*
Validation	27.3%	*(13.3-45.5)*	13.3%	*(8.5-19.4%)*	20.1%	*(15.5-25.3%)*	13.5%	*(5.6-25.8%)*	14.6%	*(8.4-22.9%)*
Negative Predictive Value (%)										
Training	99.2%	*(98.0-99.8%)*	98.4%	*(97.3-99.2%)*	96.3%	*(94.5-97.6%)*	99.6%	*(99.1-99.9%)*	99.3%	*(98.5-99.7%)*
Validation	99.6%	*(97.7-100%)*	98.9%	*(97.1-99.7%)*	95.7%	*(92.8-97.7%)*	98.9%	*(97.5-99.6%)*	99.4%	*(98.2-99.9%)*

Table 5. Predictive statistics for spontaneous preterm birth (sPTB) at less than 30, 34 and 37 weeks' gestation and within 1 and 2 weeks of testing in the group of women with transvaginal USS cervical length (CL) by training and validation sets.

CL group	**sPTB at less than**						**sPTB within**			
Outcome	**30 wk**	***95% CI***	**34 wk**	***95% CI***	**37 wk**	***95% CI***	**1 wk**	***95% CI***	**2 wk**	***95% CI***
Sensitivity %										
Training	94.1%	*(71.3-99.9)*	92.7%	*(80.1-98.5%)*	100%	*(94.8-100%)*	87.5%	*(47.3-99.7%)*	81.0%	*(58.1-94.6%)*
Validation	88.9%	*(51.8-99.7%)*	100%	*(80.5-100%)*	100%	*(89.1-100%)*	57.1%	*(18.4-90.1%)*	75.0%	*(34.9-96.8%)*
Specificity %										
Training	63.8%	*(55.0-72.1)*	35.8%	*(28.7-43.5%)*	8.1%	*(4.4-13.5%)*	81.0%	*(75.2-85.9%)*	66.8%	*(60.0-73.2%)*
Validation	61.4%	*(50.1-71.9%)*	34.6%	*(26.6-43.3%)*	5.7%	*(2.3-11.4%)*	78.4%	*(70.9-84.7%)*	63.3%	*(54.9-71.1%)*
Balanced accuracy *((Sens.+Spec)/2)*										
Training	78.98%	*(69.94-85.86%)*	64.26%	*(56.75-71.13%)*	54.06%	*(47.19-60.78%)*	84.25%	*(68.68-92.88%)*	73.89%	*(63.78-81.97%)*
Validation	75.17%	*(60.75-85.55%)*	67.29%	*(57.50-75.78%)*	52.85%	*(43.43-62.06%)*	67.76%	*(46.69-83.45%)*	69.13%	*(51.77-82.37%)*
Likelihood ratio - positive										
Training	2.60	*(2.01-3.37)*	1.44	*(1.25-1.66)*	1.09	*(1.04-1.14)*	4.60	*(3.16-6.72)*	2.44	*(1.84-3.24)*
Validation	2.31	*(1.61-3.29)*	1.53	*(1.35-1.73)*	1.06	*(1.02-1.11)*	2.64	*(1.30-5.38)*	2.04	*(1.30-3.21)*
Likelihood ratio - negative										
Training	0.09	*(0.01-0.62)*	0.20	*(0.07-0.62)*	0.00	-	0.15	*(0.02-0.97)*	0.29	*(0.12-0.69)*
Validation	0.18	*(0.03-1.16)*	0.00	-	0.00	-	0.55	*(0.23-1.29)*	0.40	*(0.12-1.32)*
Positive Predictive Value (%)										
Training	25.4%	*(15.3-37.9%)*	25.5%	*(18.7-33.3%)*	31.9%	*(25.8-38.6%)*	14.3%	*(5.9-27.2%)*	19.8%	*(12.0-29.8%)*
Validation	20.0%	*(9.1-35.6%)*	16.3%	*(9.8-24.9%)*	21.6%	*(15.3-29.1%)*	11.1%	*(3.1-26.1%)*	10.0%	*(3.8-20.5%)*
Negative Predictive Value (%)										
Training	98.8%	*(93.5-100%)*	95.4%	*(87.1-99.0%)*	100%	*(75.3-100%)*	99.4%	*(96.9-100%)*	97.2%	*(93.0-99.2%)*
Validation	98.1%	*(89.7-100%)*	100%	*(92.3-100%)*	100%	*(59.0-100%)*	97.5%	*(92.8-99.5%)*	97.9%	*(92.6-99.7%)*

Table 6. Predictive statistics for spontaneous preterm birth (sPTB) at less than 30, 34 and 37 weeks' gestation and within 1 and 2 weeks of testing in the group of women with both cervicovaginal fluid fetal fibronectin (qfFN) and transvaginal USS cervical length (CL) by training and validation sets.

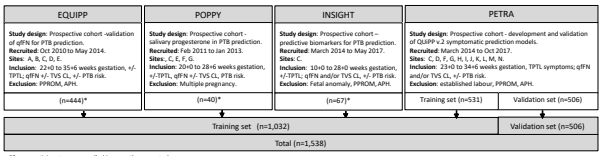
qfFN & CL group		**sPTB at less than**						**sPTB within**			
Outcome		**30 wk**	***95% CI***	**34 wk**	***95% CI***	**37 wk**	***95% CI***	**1 wk**	***95% CI***	**2 wk**	***95% CI***
	Sensitivity %										
Training		100%	*(80.5-100%)*	92.7%	*(80.1-98.5%)*	94.2%	*(85.8-98.4%)*	100%	*(63.1-100%)*	90.5%	*(69.6-98.8%)*
Validation		100%	*(63.1-100%)*	93.8%	*(69.8-99.8%)*	100%	*(88.8-100%)*	85.7%	*(42.1-99.6%)*	100%	*(63.1-100%)*
	Specificity %										
Training		68.5%	*(59.7-76.3%)*	44.5%	*(37.0-52.2%)*	16.3%	*(10.9-22.9%)*	78.7%	*(72.7-83.9%)*	69.2%	*(62.5-75.4%)*
Validation		60.0%	*(48.0-71.1%)*	39.3%	*(30.6-48.6%)*	16.1%	*(9.8-24.2%)*	75.7%	*(67.6-82.7%)*	60.0%	*(51.2-68.3%)*
	Balanced Accuracy *((Sens.+Spec)/2)*										
Training		84.23%	*(77.72-89.11%)*	68.60%	*(61.40-74.99%)*	55.23%	*(48.37-61.89%)*	89.37%	*(83.77-91.19%)*	79.85%	*(71.41-86.28%)*
Validation		80.00%	*(69.22-87.68%)*	66.55%	*(55.66-75.91%)*	58.04%	*(48.82-66.72%)*	80.72%	*(63.55-90.96%)*	80.00%	*(70.46-87.03%)*
	Likelihood ratio - positive										
Training		3.17	*(2.46-4.08)*	1.67	*(1.43-1.96)*	1.12	*(1.03-1.23)*	4.70	*(3.65-6.06)*	2.94	*(2.30-3.76)*
Validation		2.50	*(1.89-3.30)*	1.55	*(1.28-1.87)*	1.19	*(1.10-1.29)*	3.53	*(2.31-5.40)*	2.50	*(2.03-3.07)*
	Likelihood ratio - negative										
Training		0.00	-	0.16	*(0.05-0.49)*	0.36	*(0.13-0.98)*	0.00	-	0.14	*(0.04-0.52)*
Validation		0.00	-	0.16	*(0.02-1.07)*	0.00	-	0.19	*(0.03-1.16)*	0.00	-
	Positive Predictive Value (%)										
Training		29.3%	*(18.1-42.7%)*	28.4%	*(20.9-36.8%)*	32.7%	*(26.2-39.7%)*	14.5%	*(6.5-26.7%)*	22.9%	*(14.4-33.4%)*
Validation		21.1%	*(9.6-37.3%)*	16.9%	*(9.8-26.3%)*	24.8%	*(17.5-33.3%)*	15.4%	*(5.9-30.5%)*	12.9%	*(5.7-23.9%)*
	Negative Predictive Value (%)										
Training		100%	*(95.9-100%)*	96.3%	*(89.4-99.2%)*	86.7%	*(69.3-96.2%)*	100%	*(97.9-100%)*	98.6%	*(95.1-99.8%)*
Validation		100%	*(92.1-100%)*	98.0%	*(89.1-99.9%)*	100%	*(81.5-100%)*	99.0%	*(94.8-100%)*	100%	*(95.5-100%)*

Table 7. Comparison of area under the ROC curve (AUC) between QUiPP, prediction of spontaneous preterm birth (sPTB) at less than 30, 34 and 37 weeks' gestation, and at 1 and 2 weeks post test, using fetal fibronectin (qfFN) test alone, cervical length (CL) alone, and both tests combined.

AUC for prediction of sPTB < 30 weeks (n=83*)				
	AUC	Std Err	*95%CI*	*Pr>chi² ***
Both qfFN & CL	0.953	0.028	*(0.899-1.000)*	*standard*
qfFN alone	0.907	0.038	*(0.832-0.982)*	*0.14*
CL alone	0.848	0.079	*(0.693-1.000)*	*0.17*
AUC for prediction of sPTB < 34 weeks (n=138*)				
	AUC	Std Err	*95%CI*	*Pr>chi² ***
Both qfFN & CL	0.831	0.052	*(0.729-0.933)*	*standard*
qfFN alone	0.783	0.062	*(0.662-0.905)*	*0.09*
CL alone	0.789	0.055	*(0.683-0.896)*	*0.35*
AUC for prediction of sPTB < 37 weeks (n=143)				
	AUC	Std Err	*95%CI*	*Pr>chi²***
Both qfFN & CL	0.731	0.051	*(0.630-0.831)*	*standard*
qfFN alone	0.692	0.053	*(0.589-0.796)*	*0.24*
CL alone	0.719	0.050	*(0.619-0.819)*	*0.75*
AUC for prediction of sPTB < 1 week (n=143)				
	AUC	Std Err	*95%CI*	*Pr>chi²***
Both qfFN & CL	0.875	0.056	*(0.766-0.984)*	*standard*
qfFN alone	0.893	0.042	*(0.811-0.975)*	*0.65*
CL alone	0.698	0.126	*(0.450-0.945)*	*0.01*
AUC for prediction of sPTB < 2 weeks (n=143)				
	AUC	Std Err	*95%CI*	*Pr>chi²***
Both qfFN & CL	0.889	0.049	*(0.793-0.985)*	*standard*
qfFN alone	0.904	0.036	*(0.833-0.975)*	*0.67*
CL alone	0.731	0.113	*(0.510-0.951)*	*0.02*

*number of observations. Some women were recruited at later gestations.

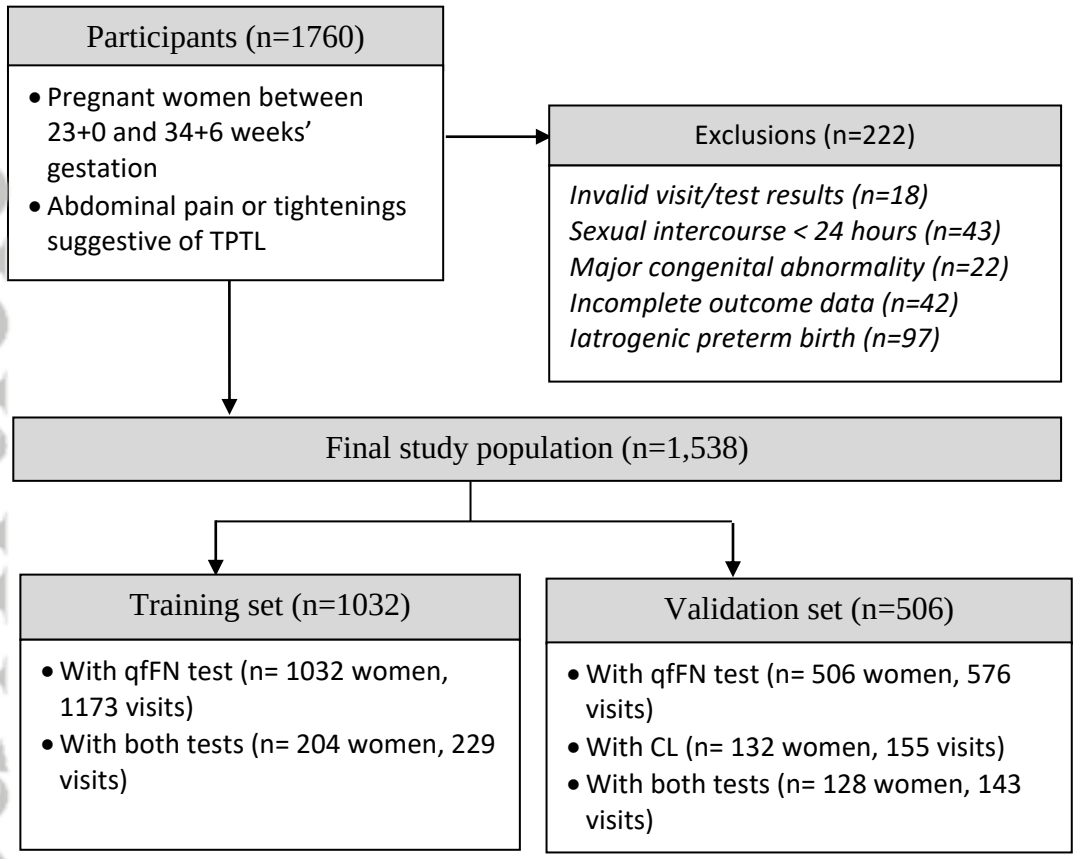
** Pr>chi2 test of significance

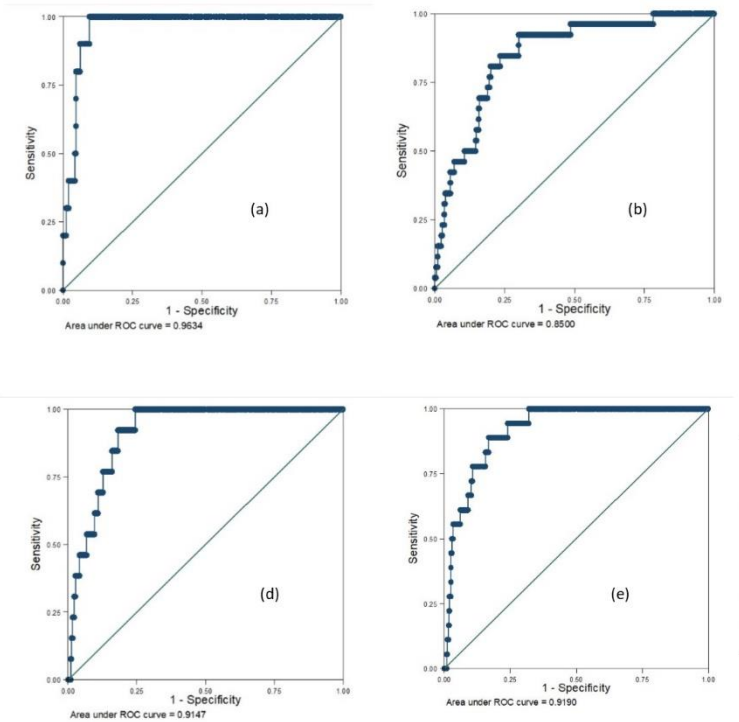


*Some participants were enrolled in more than one study.

Sites: A=tertiary hospital, Scotland; B=tertiary hospital, London; C=tertiary hospital, London; D=tertiary hospital, London; E=district general hospital, London, F= district general hospital, North West England; G= tertiary hospital, South West England; H=district general hospital, South East England; I=district general hospital, London; J=district general hospital, North East England; K=district general hospital, North West England; L=tertiary hospital, North East England; M=district general hospital, North East England; N= tertiary hospital, South West England.

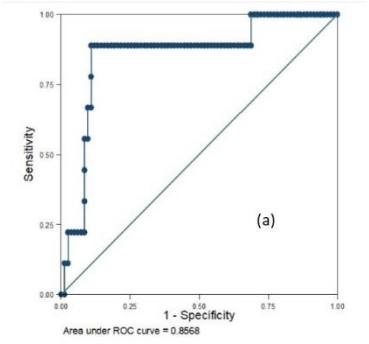
Participants were only included if symptomatic for TPTL and gestation between 23+0 and 34+6; qfFN = quantitative fetal fibronectin; PTB =preterm birth; TPTL = symptoms of threatened preterm labour; TVS CL = transvaginal cervical length; PPROM = preterm prelabour ruptured membranes, APH = antepartum haemorrhage; PTB risk =history of spontaneous preterm birth or preterm prelabour ruptured membranes, late miscarriage, cervical surgery.

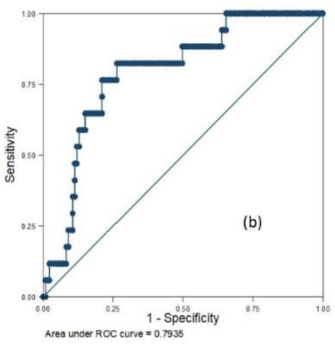


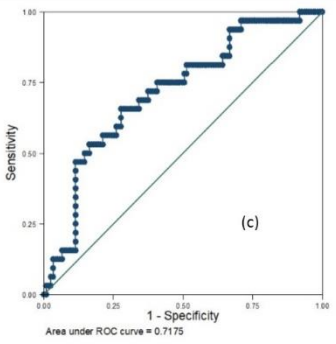


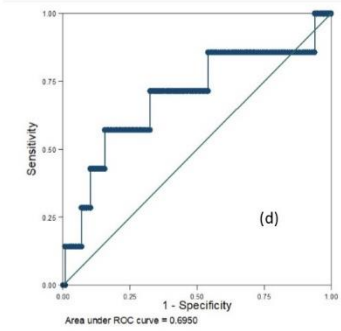
fFN group

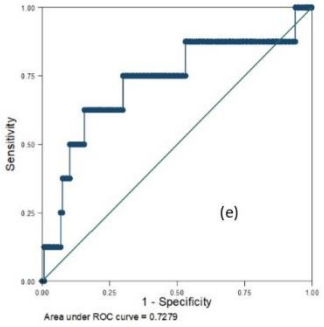
qfFN test group				
Prediction of sPTB	n	AUC	Std Err	*95%CI*
(a) <30 weeks	272	0.96	0.013	*(0.94-0.99)*
(b) <34 weeks	520	0.85	0.035	*(0.78-0.92)*
(c) <37 weeks	576	0.77	0.030	*(0.71-0.83)*
(d) <1 week	576	0.92	0.022	*(0.87-0.96)*
(e) <2 weeks	576	0.92	0.022	*(0.88-0.96)*





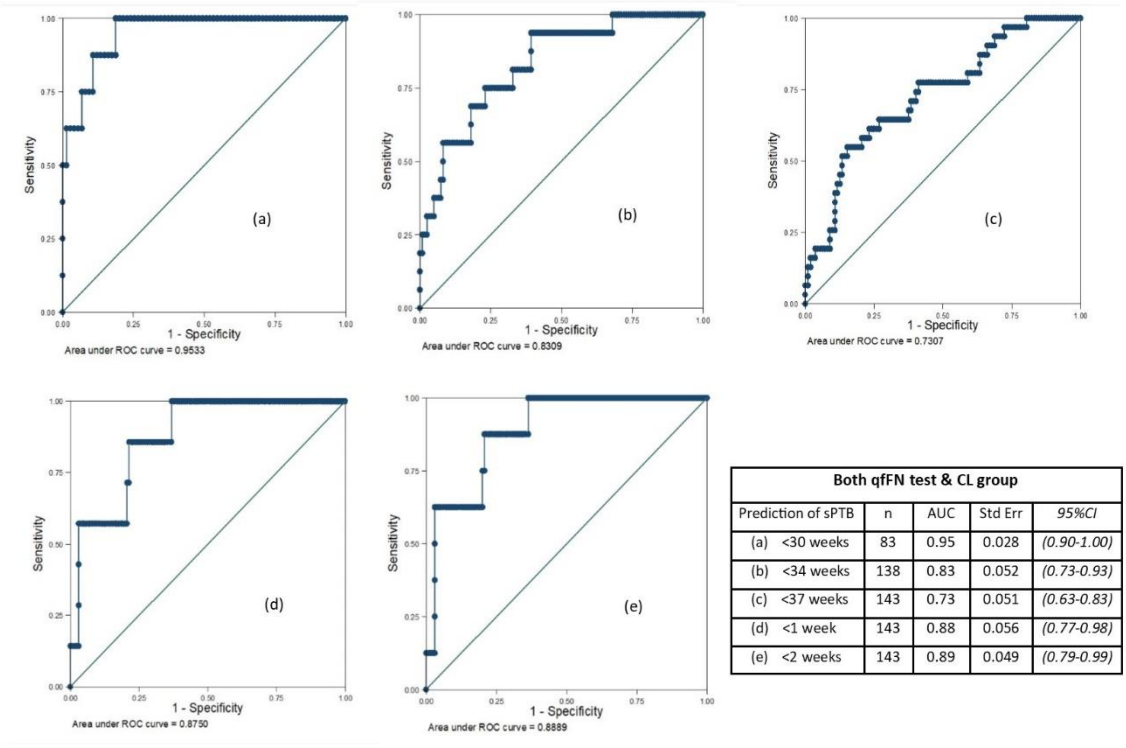






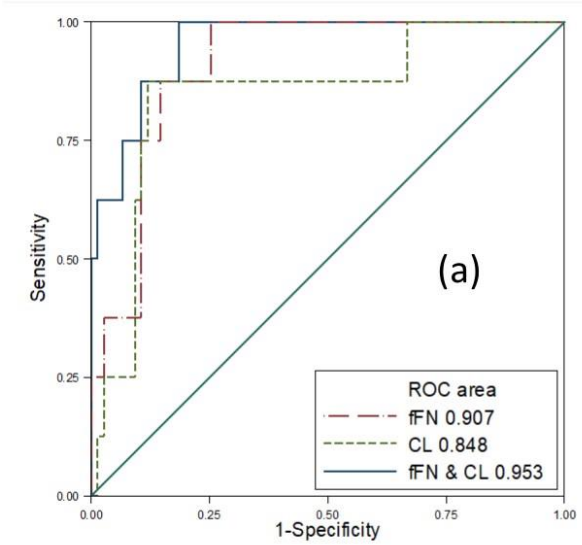
CL group

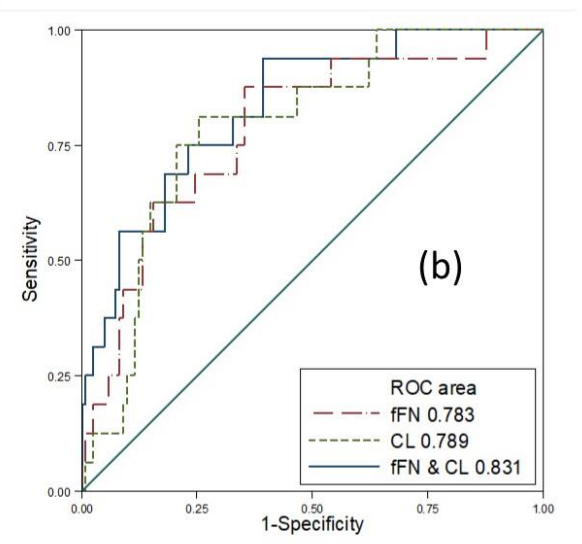
CL group				
Prediction of sPTB	n	AUC	Std Err	*95%CI*
(a) <30 weeks	92	0.86	0.073	*(0.71-0.99)*
(b) <34 weeks	150	0.79	0.053	*(0.69-0.90)*
(c) <37 weeks	155	0.72	0.050	*(0.62-0.82)*
(d) <1 week	155	0.70	0.127	*(0.45-0.95)*
(e) <2 weeks	155	0.73	0.114	*(0.51-0.95)*

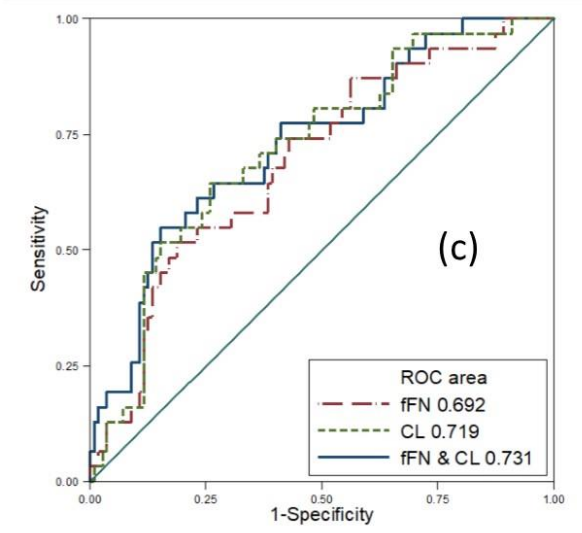


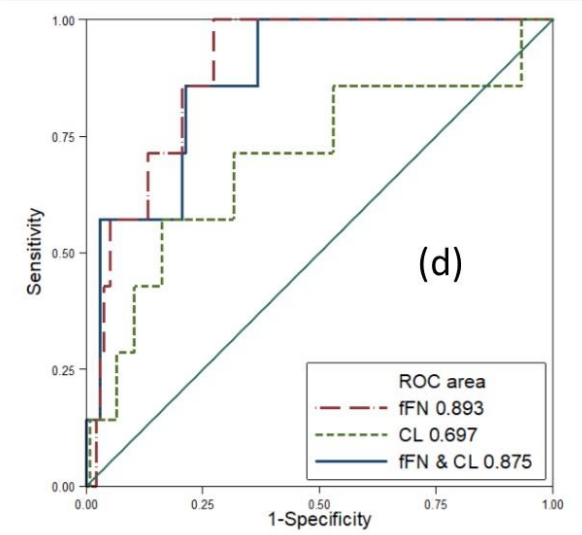
fFN+CL group

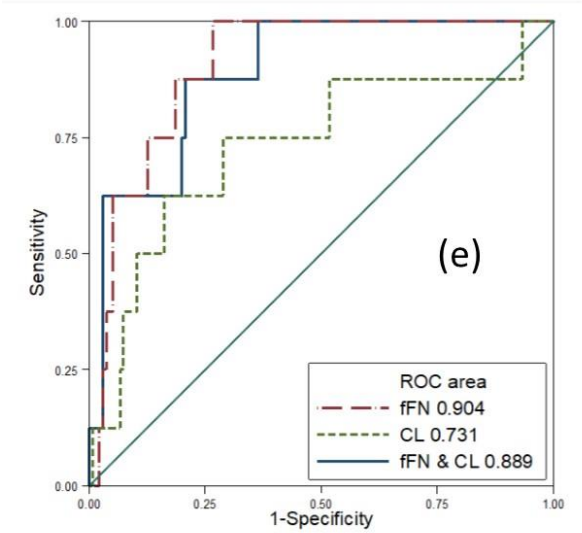
Both qfFN test & CL group				
Prediction of sPTB	n	AUC	Std Err	*95%CI*
(a) <30 weeks	83	0.95	0.028	*(0.90-1.00)*
(b) <34 weeks	138	0.83	0.052	*(0.73-0.93)*
(c) <37 weeks	143	0.73	0.051	*(0.63-0.83)*
(d) <1 week	143	0.88	0.056	*(0.77-0.98)*
(e) <2 weeks	143	0.89	0.049	*(0.79-0.99)*











	Prediction of sPTB at:
a	< 30 weeks
b	< 34weeks
c	< 37 weeks
d	≤ 1 week
e	≤ 2 weeks