Verification of Placental Growth Factor and Soluble-fms-like Tyrosine Kinase 1 Assay Performance in Late Pregnancy and their Diagnostic Test Accuracy in Women with Reduced Fetal Movement

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Title: Verification of Placental Growth Factor and Soluble-fms-like Tyrosine Kinase 1 Assay Performance in Late Pregnancy and their Diagnostic Test Accuracy in Women with Reduced Fetal Movement

Authors: Lindsay K. Armstrong-Buisseret[1], Shonagh Haslam[2], Tim James[3], Lucy Bradshaw[1] and Alexander E.P. Heazell[4,5]

[1]Nottingham Clinical Trials Unit, University of Nottingham

[2]Department of Clinical Biochemistry, Royal Preston Hospital

[3]Department of Clinical Biochemistry, John Radcliffe Hospital

[4]Maternal and Fetal Health Research Centre, University of Manchester

[5]St. Mary's Hospital, Central Manchester University Hospitals NHS Foundation Trust

Corresponding author: Dr Lindsay Armstrong-Buisseret, Nottingham Clinical Trials Unit, Building 42, Room A17, University of Nottingham, University Park, Nottingham NG7 2RD.

Email: Lindsay.Armstrong-Buisseret@nottingham.ac.uk

Word count (abstract): 250

Word count (excluding abstract and references): 4639

Abstract

Background: Placental growth factor (PlGF) and soluble-fms-like tyrosine kinase 1 (sFlt-1) are biomarkers of placental function used to aid diagnosis and prediction of pregnancy complications. This work verified the analytical performance of both biomarkers and provides preliminary diagnostic accuracy data to identify adverse pregnancy outcome in women with reduced fetal movement (RFM).

Methods: Verification of sFlt-1 and PlGF assays included a comparative accuracy assessment of 24 serum samples analysed at six different sites, and laboratory specific precision estimates. The sFlt-1/PlGF ratio was assessed in serum samples obtained prospectively from 295 women with RFM ≥36 weeks' gestation; diagnostic accuracy was evaluated using 2x2 tables and Area Under the Receiver Operator Characteristic (AUROC) curve.

Results: Regression analysis showed performance between sites was good with Passing-Bablok slopes ranging from 0.96 to 1.05 (sFlt-1) and 0.93 to 1.08 (PlGF). All sites had a mean bias <15% although there was poorer agreement at the lowest PlGF concentrations. All within and between batch coefficients of variation were <10%. In 289 women with an appropriately grown fetus, a sFlt-1/PlGF ratio ≥38 had a sensitivity of 0.20 (95% confidence interval [CI] 0.07, 0.41), specificity of 0.88 (95% CI 0.83, 0.92) and AUROC curve of 0.58 (95% CI 0.47, 0.68) to identify adverse pregnancy outcome.

Conclusions: Analytical performance of the sFlt-1 and PlGF assays was comparable across different sites. The sensitivity of sFlt-1/PlGF to identify adverse pregnancy outcome in women with RFM was considered acceptable, in the absence of other tests, to progress to a pilot randomised controlled trial.

Introduction

Placental growth factor (PlGF) and soluble fms-like tyrosine kinase-1 (sFlt-1) are proteins present in maternal circulation that hold significant promise as biomarkers to aid the diagnosis and management of pregnancy complications, particularly those that relate to placental dysfunction.[1-5] Research evaluating these biomarkers initially focused on detection and management of preeclampsia[4] but more recent studies suggest they have potential value in identifying other complications including: prediction of small for gestational age (SGA) infants,[5, 6] fetal compromise after maternal presentation with reduced fetal movement (RFM)[7] and the need for intervention during labour, and compromise at the time of birth.[8] A systematic review of diagnostic test accuracy (DTA) studies found that PlGF was the most accurate biochemical test to identify pregnancies that would end in stillbirth with a diagnostic odds ratio of 49.2 (95% confidence interval [CI] 12.7, 191), however this information comes from 5,894 individuals, of whom only 16 had a stillbirth.[9] While this is an important avenue of exploration to prevent stillbirth, further studies are needed to increase understanding of test accuracy of PlGF in the prediction of adverse pregnancy outcome.

Measurement of sFlt-1 and PlGF utilises immunoassays, the majority of which are automated. While there are a number of reports describing the diagnostic accuracy of these tests in a variety of clinical settings,[3, 5, 10-12] descriptions of method performance are relatively limited. It is important to conduct such studies since the utility of these assays in clinical practice depends upon reproducible, accurate analysis. In addition, introduction of protocols using common thresholds requires an understanding of how analytical methods compare between laboratories. PlGF is bound in maternal blood by sFlt-1 and assays measure either unbound PlGF or the ratio of sFlt-1 to PlGF with the latter performing as well as, or better than, each biomarker alone in the diagnosis of preeclampsia.[3, 13, 14] Results for the Elecsys® sFlt-1 and Elecsys® PlGF

immunoassays (Roche Diagnostics; Germany) are comparable across a number of sites[15] and similar analytical performance of these biomarkers has been demonstrated using other platforms including BRAHMS® Kryptor (ThermoFisher Scientific; Germany),[16] Alere Triage® PlGF test (Alere Inc; USA), DELFIA® Xpress PlGF 1-2-3 test (PerkinElmer; Finland) and ELISA (R&D Systems; USA).[17, 18] However, the methods have undergone further refinement and there is limited information regarding current performance including susceptibility to biotin interference which is a potential issue for some immunoassay systems.[19-21]

To understand method robustness in a wider clinical and laboratory context, we conducted a verification of analytical performance for the Elecsys® sFlt-1 and Elecsys® PlGF assays across sites participating in the ReMIT-2 trial.[22] This is a multicentre, randomised controlled, interventional pilot trial investigating whether including measurement of the sFlt-1/PlGF ratio following maternal presentation with RFM ≥36 weeks' gestation with the aim of preventing adverse pregnancy outcome is feasible. This population was chosen for this study because RFM is associated with increased risk of adverse pregnancy outcomes and placental dysfunction,[23-25] but management strategies based upon ultrasound scanning and induction of labour for RFM in all cases are associated with increased rates of obstetric intervention.[26] Our prior data indicating improved test performance with inclusion of PlGF assessment would potentially allow intervention to be focussed on women with RFM who were more likely to benefit from early induction.[7] Here we report the performance of sFlt-1 and PlGF across six UK laboratory sites and assess potential confounding pre-analytical factors including temperature stability and biotin interference. In addition, preliminary DTA data in this clinical context are presented, the objective of which was to evaluate the diagnostic accuracy of the sFlt-1/PlGF ratio in

identifying a composite adverse pregnancy outcome in women presenting with RFM ≥ 36 weeks' gestation.

Materials and Methods

Preparation, Distribution and Analysis of Samples

Blood samples were collected by venepuncture into clot activator gel tubes (Sarstedt; Nümbrecht, Germany), centrifuged at 3000 g for 10 min, and the resultant serum split into aliquots and stored at -80^0C. Frozen samples were sent on dry ice by same day courier and stored at -80^0C at the receiving site until required. Sites were instructed to thaw samples immediately prior to analysis. All samples were analysed using the Elecsys® sFlt-1 and Elecsys® PlGF fully automated immunoassays on three different models of Roche Cobas® analyser (801, 602 and e411; Roche Diagnostics; Mannheim, Germany). Assays were run according to manufacturer's instructions and the sFlt-1/PlGF ratio was calculated for all samples.

Verification of sFlt-1 and PlGF Analytical Performance

Currently, there are no international reference materials to provide traceability of calibration of sFlt-1 or PlGF which contributes to a lack of gold standard analytical techniques to assess the accuracy of methods. Therefore, relative accuracy was addressed using anonymised residual serum blood samples routinely received in the Clinical Biochemistry Laboratory at the Royal Preston Hospital. For the study, 24 anonymised patient serum samples were distributed for analysis across six ReMIT-2 trial sites (Royal Preston Hospital, John Radcliffe Hospital, Sunderland Royal Hospital, St George's Hospital, Royal Liverpool University Hospital and University Hospital of North Tees). The means of sFlt-1 and PlGF results for each sample were calculated and comparisons performed using Passing-Bablok regression analysis against the

mean as a target. Additionally, the percent bias was determined for each site using the mean as the comparator. The clinical impact of any bias seen between sites was assessed by comparing the sFlt-1/PlGF ratios obtained from each sample against published thresholds for pre-eclampsia[3, 10, 27] with the mean value use as the comparator. These thresholds have been adapted for clinical use by the authors as follows: the risk of developing pre-eclampsia is considered low with a sFlt-1/PlGF ratio <38 (i.e. can be ruled-out for 1 week), intermediate with a ratio ≥38 to <85 (i.e. can be ruled-in within the next 4 weeks) and high with a ratio ≥85 (i.e. is a confirmed diagnosis).

Within batch precision was determined on the 801 analyser at one site (Royal Preston Hospital) using three serum samples ($n = 10$). Between batch precision was assessed using two levels of internal quality control (IQC) materials (PreciControl Multimarker, Roche Diagnostics; Mannheim, Germany) on the Cobas® analyser at all eight ReMIT-2 sites (Manchester Royal Infirmary and James Cook University Hospital in addition to those listed above). Between batch precision data was collected to reflect real live usage over a time period to reach $n = 20$. The timeframe was dependent on the numbers done at each site and varied between three weeks and three months. For both assays, the lower limit of quantitation (LLOQ) was defined as the lowest measureable concentration with an acceptable precision and was assessed by analysing three low concentration samples in 10 separate runs. For both precision and LLOQ studies the mean, standard deviation (SD) and coefficient of variation (CV) were calculated.

The stability of sFlt-1 and PlGF was tested in five serum samples at room temperature over 24h, at 4°C over 4 weeks, at -20°C over 6 months and over 3 freeze thaw cycles. Biotin interference in the sFlt-1 and PlGF assays was determined at concentrations of 30, 60 and 500 ng/mL biotin. These were chosen to reflect serum biotin concentrations which might occur

following ingestion of over-the-counter supplements or in patients taking mega doses for treating progressive Multiple Sclerosis.[21, 28] Two serum pools were spiked with increasing doses of biotin (Sigma-Aldrich; Gillingham, UK) according to methods described previously.[29] Working solutions were made up in distilled water and the spiked amount was no more than 5% of the total serum volume. Additionally, distilled water was spiked as a control sample to account for dilutional or matrix effects. The mean percentage difference between the control sample and the spiked serum pools were then calculated.

Diagnostic Test Accuracy Study

A preliminary DTA study was conducted to evaluate the sFlt-1/PlGF ratio (as the index test) for identifying adverse pregnancy outcome using maternal serum samples previously collected in three consecutive, separate studies run in a single tertiary UK maternity unit.[7, 30] Each of these had ethics approval for measurement of placental analytes (FEMINA Oldham Research Ethics Committee [REC] reference 08/H1011/83, recruitment took place from August 2009 – October 2019; FEMINA2 Greater Manchester West REC reference 11/NW/0650, recruitment took place from January 2012 – May 2014; FEMINA3 Greater Manchester East REC reference 16/NW/0481, recruitment started in August 2016 and is still ongoing) and written consent was obtained from all participants. Relevant eligibility criteria included women with a viable singleton pregnancy presenting with a primary complaint of maternal perception of RFM ≥36 weeks' gestation and having had an ultrasound scan for fetal biometry, liquor volume and umbilical artery Doppler on presentation. Exclusion criteria included maternal age <18 years and fetus with a known congenital abnormality. Samples were analysed at site 6 using a Cobas® e411 and as this took place after birth, clinical care was not influenced by the sFlt-1/PlGF results, and assay results were not available when classifying pregnancy outcome.

Clinical and demographic details were collected from participants at presentation, along with ultrasound scan results for fetal biometry, liquor volume and umbilical artery Doppler. Delivery and pregnancy outcomes included gestational age at delivery, birthweight and gender. The target condition was a composite adverse pregnancy outcome defined as birthweight <5th centile (derived from individualised birth weight centiles[31]), neonatal intensive care unit (NICU) admission, umbilical artery pH <7.1 at birth or stillbirth/neonatal death. Admission to NICU was regarded as any admission within the neonatal period regardless of indication. Where umbilical artery pH was missing it was assumed to be normal since this would not be measured in normal healthy babies. The reference standards were each of the four individual components of the composite outcome. Since stillbirth alone as a measure of adverse pregnancy outcome would require a very large number of participants, a composite outcome was used similar to those employed in other perinatal trials.[32, 33] As sFlt-1 and PlGF were not measured until after birth, the clinicians determining whether the reference standard was present were not aware of the index test results. Similarly, those measuring the index test were unaware of whether the reference standard was present.

We aimed to include 410 participants in the DTA study to achieve a margin of error of 10% for sensitivity and 2.3% for specificity assuming that the prevalence of adverse pregnancy outcome would be 10%, sensitivity of the sFlt-1/PlGF ratio would be 80% and specificity would be 95%. The diagnostic accuracy of sFlt-1, PlGF and the sFlt-1/PlGF ratio was calculated using Stata Version 15 (StataCorp, College Station, USA). Performance of the sFlt-1/PlGF ratio was assessed using predefined levels of 38 (determined to be the optimal level to diagnose preeclampsia[3]) and 112 (95th centile value >37 weeks' gestation). The diagnostic accuracy of currently used tests including estimated fetal weight (EFW) <10th centile, and oligohydramnios (defined by gestation-specific centile for amniotic fluid index or maximal

pool depth) as well as EFW $<10^{th}$ centile or sFlt-1/PlGF ratio ≥ 38 were also examined. Sensitivity, specificity, and positive and negative predictive values were calculated along with the Area Under the Receiver Operator Characteristic (AUROC) curve for the sFlt-1/PlGF ratio; 95% CIs were calculated for test performance characteristics. Any instances of indeterminate or missing sFlt-1/PlGF ratios, or where all components for the adverse pregnancy outcome were indeterminate or missing, were excluded from the analysis.

Results

### *Verification of sFlt-1 and PlGF Analytical Performance*

Results for sFlt-1, PlGF and the sFlt-1/PlGF ratio from 24 samples analysed at each site were compared using Passing-Bablok analysis with the mean of all sites as the comparator. For all sites and assays, this gave slopes ranging from 0.89 to 1.08 and intercepts ranging from -6.52 to 9.13 pg/mL (Table 1). In addition, percent bias for sFlt-1, PlGF and the sFlt-1/PlGF ratio from each site was plotted against the mean of all results which showed that for both biomarkers, most sites had a mean bias of <10% (Figure 1). The only exception was laboratory 6 which had a mean bias for PlGF of -10.2% (Figure 1b), primarily due to greater variability at concentrations below 25 pg/mL when bias values ranged from -56 to 24% (Figure 1b). The variability at low PlGF concentrations impacted on the bias of the calculated sFlt-1/PlGF ratio; however over 70% of the sFlt-1/PlGF ratio results had a bias <10% (Figure 1c). Assessment of the clinical impact of this bias performed on individual sFlt-1/PlGF ratio results (n = 144) indicated that two samples (1.4%) would have been classified as having an intermediate risk of developing pre-eclampsia where the consensus mean was low risk.

Table 1: Passing-Bablok slope and intercept results for sFlt-1, PlGF and sFlt-1/PlGF ratio at each site

Site	Slope			Intercept		
	sFlt-1 (95% CI)	PlGF (95% CI)	sFlt-1/PlGF ratio (95% CI)	sFlt-1 (95% CI)	PlGF (95% CI)	sFlt-1/PlGF ratio (95% CI)
1	0.96 (0.93, 0.96)	1.00 (0.99, 1.03)	0.89 (0.81, 0.92)	-2.29 (-4.84, 3.38)	2.15 (1.33, 3.14)	0.08 (-0.28, 0.56)
2	1.00 (1.00, 1.02)	0.97 (0.95, 0.98)	1.04 (1.02, 1.07)	-3.82 (-7.14, 0.42)	0.22 (-0.44, 0.75)	-0.17 (-0.53, 0.06)
3	1.05 (1.04, 1.06)	1.03 (1.02, 1.07)	0.97 (0.94, 0.97)	-6.52 (-11.59, -1.18)	0.72 (-0.26, 1.09)	-0.02 (-0.26, 0.19)
4	0.96 (0.95, 1.05)	0.93 (0.91, 0.95)	0.97 (0.92, 1.09)	9.13 (-15.16, 16.15)	1.70 (0.60, 2.23)	0.46 (-0.29, 1.26)
5	1.00 (0.99, 1.00)	0.99 (0.98, 1.00)	1.00 (0.94, 1.02)	-1.04 (-2.65, 3.18)	0.63 (-0.03, 1.05)	-0.24 (-0.52, 0.18)
6	1.03 (1.01, 1.03)	1.08 (1.07, 1.09)	1.01 (0.98, 1.41)	-2.12 (-5.45, 10.19)	-5.11 (-5.19, -4.28)	-0.10 (-2.35, 0.97)

CI: confidence interval; PlGF: placental growth factor; sFlt-1: soluble fms-like tyrosine kinase-1

Figure 1: Percent bias for each site for sFlt-1, PlGF and sFlt-1/PlGF ratio ($n = 24$)
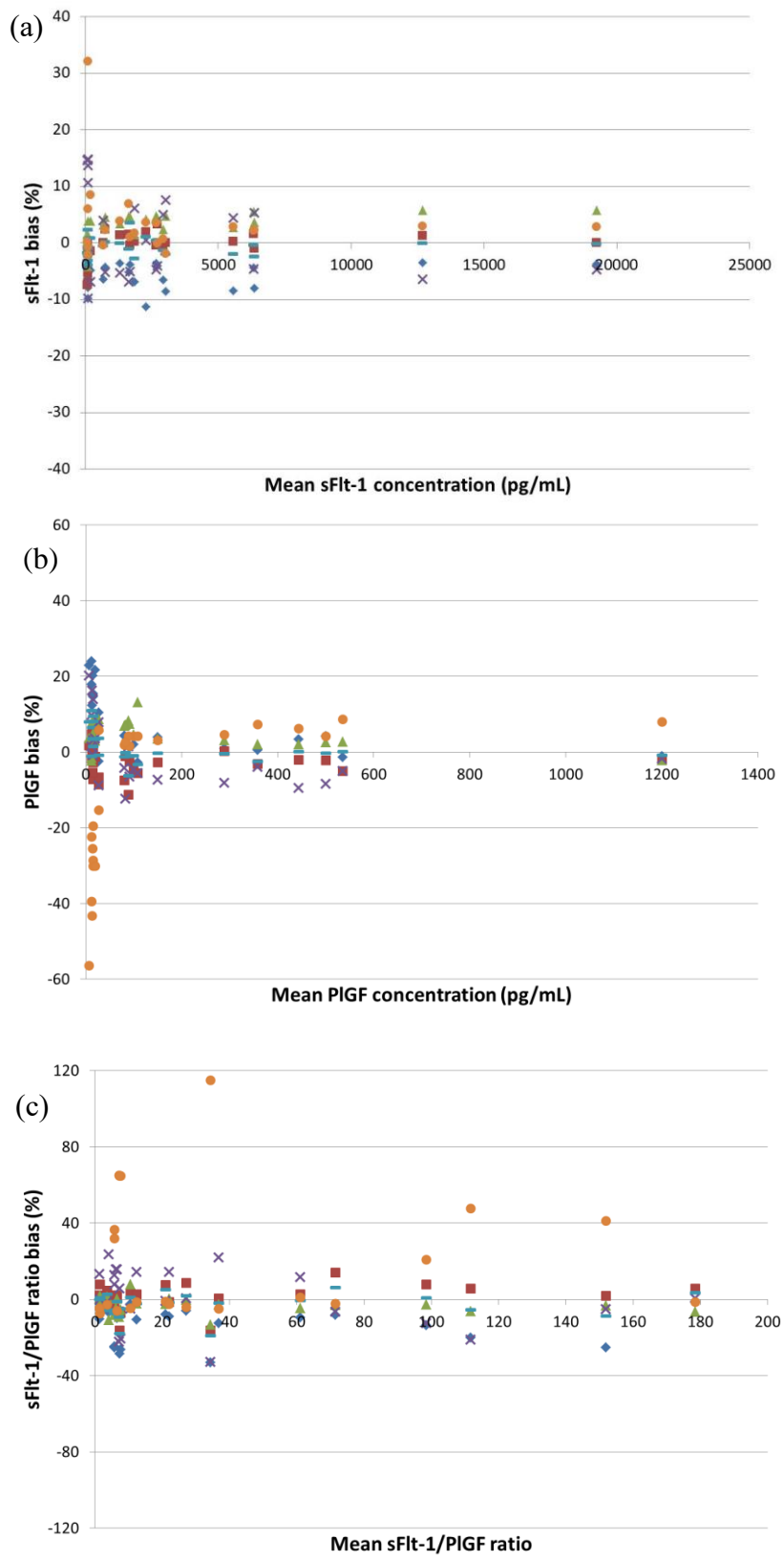
The individual site percent bias is plotted against the mean of all sites for: (a) sFlt-1; (b) PlGF; (c) sFlt-1/PlGF ratio. PlGF: placental growth factor; sFlt-1: soluble fms-like tyrosine kinase-1

◆ Site 1 (801 Cobas® analyser); ■ Site 2 (602 Cobas® analyser); ▲ Site 3 (e411 Cobas® analyser);

x Site 4 (801 Cobas® analyser); - Site 5 (602 Cobas® analyser); ● Site 6 (e411 Cobas® analyser).

Within batch precision CVs for sFlt-1 were 1.13%, 0.53% and 0.63% at 82.1 pg/mL, 1020.6 pg/mL and 13393.1 pg/mL respectively ($n = 10$). Within batch precision CVs for PlGF were 1.20%, 0.69% and 1.04% at 14.5 pg/mL, 143.9 pg/mL and 1319.3 pg/mL respectively ($n = 10$). Between batch precision estimates were generally <5% at both levels of IQC material, mean CVs for all eight sites being 4.6% (Low QC) and 4.9% (High QC) for sFlt-1 along with 4.2% (Low QC) and 4.3% (High QC) for PlGF (Table 2). Precision estimates at one site with the e411 analyser were much higher (>20% for sFlt-1 and around 10% for PlGF) although this was not instrument type related as precision at other sites with the e411 analyser was <5% for both biomarkers at the concentrations assessed.

Table 2: Between batch method precision

Test	Site (Roche Cobas® analyser)	Low QC		High QC	
		Mean (SD)	%CV	Mean (SD)	%CV
sFlt-1 (*n* = 20)	Site 1 (801)	98.1 (2.6)	2.7	959.9 (8.6)	0.9
	Site 2 (602)	110.7 (3.2)	2.9	994.3 (40.7)	4.1
	Site 3 (e411)	102.0 (21.9)	21.4	941.9 (200.7)	21.3
	Site 4 (801)	105.4 (1.8)	1.7	959.6 (40.8)	4.3
	Site 5 (602)	107.2 (1.1)	1.1	979.9 (8.8)	0.9
	Site 6 (e411)	99.1 (3.4)	3.4	986.1 (29.4)	3.0
	Site 7 (e411)	105.5 (2.0)	1.9	1020.7 (20.7)	2.0
	Site 8[a] (801)	104.6 (1.9)	1.8	938.6 (24.4)	2.6
PlGF (*n* = 20)	Site 1 (801)	108.4 (5.6)	5.2	1062.7 (54.3)	5.1
	Site 2 (602)	86.4 (3.9)	4.8	953.9 (33.4)	3.5
	Site 3 (e411)	87.5 (8.0)	9.2	936.7 (95.0)	10.1
	Site 4 (801)	86.5 (4.9)	5.6	888.9 (48.1)	5.4
	Site 5 (602)	84.7 (0.8)	1.0	939.2 (8.2)	0.9
	Site 6 (e411)	99.2 (2.6)	2.6	1072.2 (53.2)	5.0
	Site 7 (e411)	104.5 (1.7)	1.7	1041.7 (18.0)	1.7
	Site 8[b] (801)	86.6 (2.7)	3.1	930.0 (26.4)	2.8

[a]*n* = 9 for sFlt-1 low QC; *n* = 8 for sFlt-1 high QC.

[b]*n* = 7 for PlGF low QC; *n* = 10 for PlGF high QC.

CV: coefficient of variation; PlGF: placental growth factor; QC: quality control; SD: standard deviation;

sFlt-1: soluble fms-like tyrosine kinase-1

For the PlGF LLOQ, mean concentrations analysed were 5.2 pg/mL, 8.9 pg/mL and 10.7 pg/mL which gave CVs of 9.05%, 4.31% and 6.24% respectively. The LLOQ studies for sFlt-1 were performed at mean concentrations of 12.1 pg/mL, 13.3 pg/mL and 21.5 pg/mL which resulted in CVs of 7.60%, 4.24% and 4.92% respectively.

sFlt-1 and PlGF were demonstrated to be stable in serum when stored at -20^0C for 6 months (Figure 2a), when stored at 4^0C for 14 days (Figure 2b), and when stored at room temperature for 24h (Figure 2c). All serum samples tested had a mean difference from baseline of <10%. Additionally, both sFlt-1 and PlGF were stable over 3 freeze thaw cycles (Figure 2d).

Figure 2: Stability of sFlt-1 and PlGF in serum over various timeframes ($n = 5$)
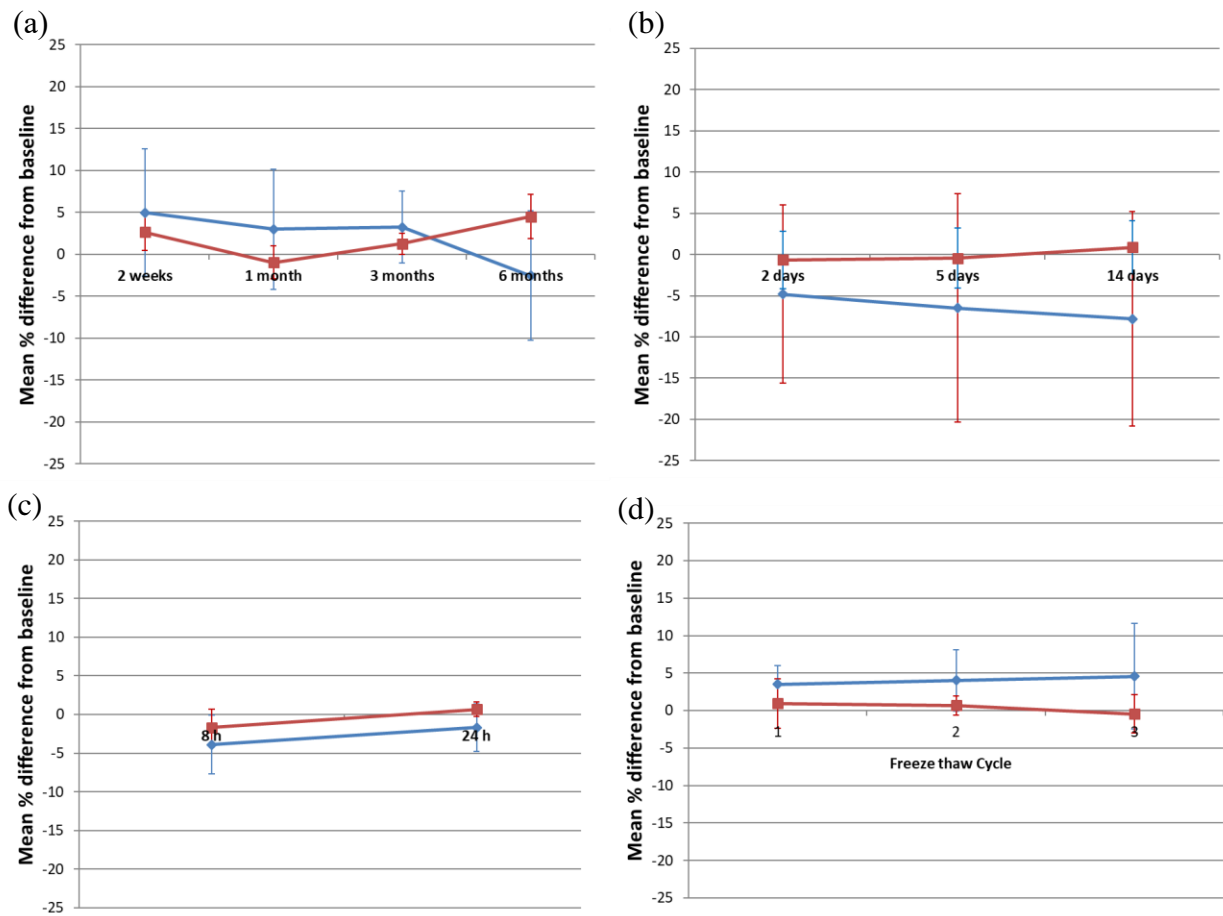
Stability of sFlt-1 and PlGF: (a) at -20^0C over 6 months; (b) at 4^0C over 14 days; (c) at room temperature over 24h; (d) over 3 freeze thaw cycles. ■ sFlt-1; ◆ PlGF; error bars are 95% confidence intervals

PlGF: placental growth factor; sFlt-1: soluble fms-like tyrosine kinase-1

Addition of biotin to pooled serum samples resulted in a reduction of >10% from the spiked control sample in the PlGF assay at biotin concentrations of 30, 60 and 500 ng/mL (Supplementary Figure 1). The effect was related to the concentration of biotin, with 500 ng/mL resulting in a greater than -80% difference. With sFlt-1, interference was not as significant with biotin concentrations of 30 and 60 ng/mL having <10% bias (Supplementary Figure 1). However, at 500 ng/mL the difference was -79% from the baseline sample.

Diagnostic Test Accuracy Study

Across the three studies, samples from 423 participants were analysed (Supplementary Figure 2). Three hundred and eighteen participants met the relevant inclusion criteria for the DTA study (≥36 weeks' gestation). Twenty three were excluded due to the EFW being <10th centile at presentation (i.e. the fetus already had a diagnosis of being SGA) since this would be an indication for delivery after 37 weeks' gestation and therefore not meet the inclusion criteria for ReMIT-2[22] and a further 6 had insufficient serum to conduct assays for both biomarkers which gave 289 participants with both sFlt-1 and PlGF results to calculate the ratio of both biomarkers. Maternal characteristics at baseline for all 318 participants are shown in Supplementary Table 1, the mean age of participants was 29.2 years (standard deviation [SD] 5.5), mean body mass index (BMI) was 26.4 (SD 5.4) and just over half (54%) were in their first pregnancy.

In 295 participants presenting with RFM ≥36 weeks' gestation whose babies were ≥10th EFW centile, 26 (8.8%) had the target condition (composite adverse pregnancy outcome); 2.4% had a birthweight <5th centile, 4.1% were admitted to NICU, 3.7% had an umbilical artery pH <7.1 at birth, and there were no stillbirths/neonatal deaths. Three babies had both low pH and NICU admission and one further baby had birthweight <5th centile and NICU admission. In babies who were ≥10th EFW centile and had results for both biomarkers ($n = 289$), the sFlt-1/PlGF ratio at a threshold of ≥38 had a sensitivity of 0.20 (95% CI 0.07, 0.41) and a specificity of 0.88 (95% CI 0.83, 0.92) to predict the composite adverse pregnancy outcome (Table 3). Applying a threshold of ≥112 for the sFlt-1/PlGF ratio had a sensitivity of 0.04 (95% CI 0.00, 0.20) with a specificity of 0.99 (95% CI 0.97, 1.00) (Table 3). The AUROC curve of the sFlt-1/PlGF ratio to identify adverse pregnancy outcomes in these women was 0.58 (95% CI 0.47, 0.68; Figure 3). For comparison, ultrasound EFW <10th centile alone ($n = 318$) had a sensitivity

of 0.24 (95% CI 0.11, 0.41) and a specificity of 0.95 (95% CI 0.91, 0.97), isolated oligohydramnios (n = 295) had a sensitivity of 0.12 (95% CI 0.02, 0.30) and a specificity of 0.85 (95% CI 0.80, 0.89) and either EFW <10th centile or sFlt-1/PlGF ≥38 (n = 312) had a sensitivity of 0.39 (95% CI 0.23, 0.58) and a specificity of 0.83 (95% CI 0.78, 0.87; Table 3) to predict the composite adverse pregnancy outcome. No significant adverse events occurred as a result of taking the blood sample to perform the sFlt-1/PlGF test, or from measuring the birthweight or umbilical artery pH.

Table 3: Diagnostic accuracy of ultrasonographic and biomarker tests to identify composite adverse pregnancy outcome

Test	Test Result	Adverse Pregnancy Outcome	Normal Pregnancy Outcome	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (95% CI)	Negative Predictive Value (95% CI)
EFW <10th centile (*n* = 318)	Positive	8	15	0.24	0.95	0.35	0.91
	Negative	26	269	(0.11, 0.41)	(0.91, 0.97)	(0.16, 0.57)	(0.87, 0.94)
Oligohydramnios in babies with EFW ≥10th centile (*n* = 295)	Positive	3	41	0.12	0.85	0.07	0.91
	Negative	23	228	(0.02, 0.30)	(0.80, 0.89)	(0.01, 0.19)	(0.87, 0.94)
Umbilical artery PI >95th centile in babies with EFW ≥10th centile (*n* = 295)	Positive	1	14	0.04	0.95	0.07	0.91
	Negative	25	255	(0.00, 0.20)	(0.91, 0.97)	(0.00, 0.32)	(0.87, 0.94)
sFlt-1/PlGF ≥38 in babies with EFW ≥10th centile (*n* = 289[a])	Positive	5	32	0.20	0.88	0.14	0.92
	Negative	20	232	(0.07, 0.41)	(0.83, 0.92)	(0.05, 0.29)	(0.88, 0.95)
sFlt-1/PlGF ≥112 in babies with EFW ≥10th centile (*n* = 289[a])	Positive	1	3	0.04	0.99	0.25	0.92
	Negative	24	261	(0.00, 0.20)	(0.97, 1.00)	(0.01, 0.81)	(0.88, 0.95)
EFW <10th centile or sFlt-1/PlGF ≥38 (n = 312[a])	Positive	13	47	0.39	0.83	0.22	0.92
	Negative	20	232	(0.23, 0.58)	(0.78, 0.87)	(0.12, 0.34)	(0.88, 0.95)

[a]Unable to calculate sFlt-1/PlGF ratio for 6 cases due to missing sFlt-1 result (*n* = 3) or PlGF result (*n* = 3), one of whom had an adverse pregnancy outcome.

CI: confidence interval; EFW: estimated fetal weight; PI: pulsatility index; PlGF: placental growth factor; sFlt-1: soluble fms-like tyrosine kinase-1

Figure 3: Receiver Operator Characteristic curve for sFlt-1/PlGF ratio to predict composite adverse pregnancy outcome
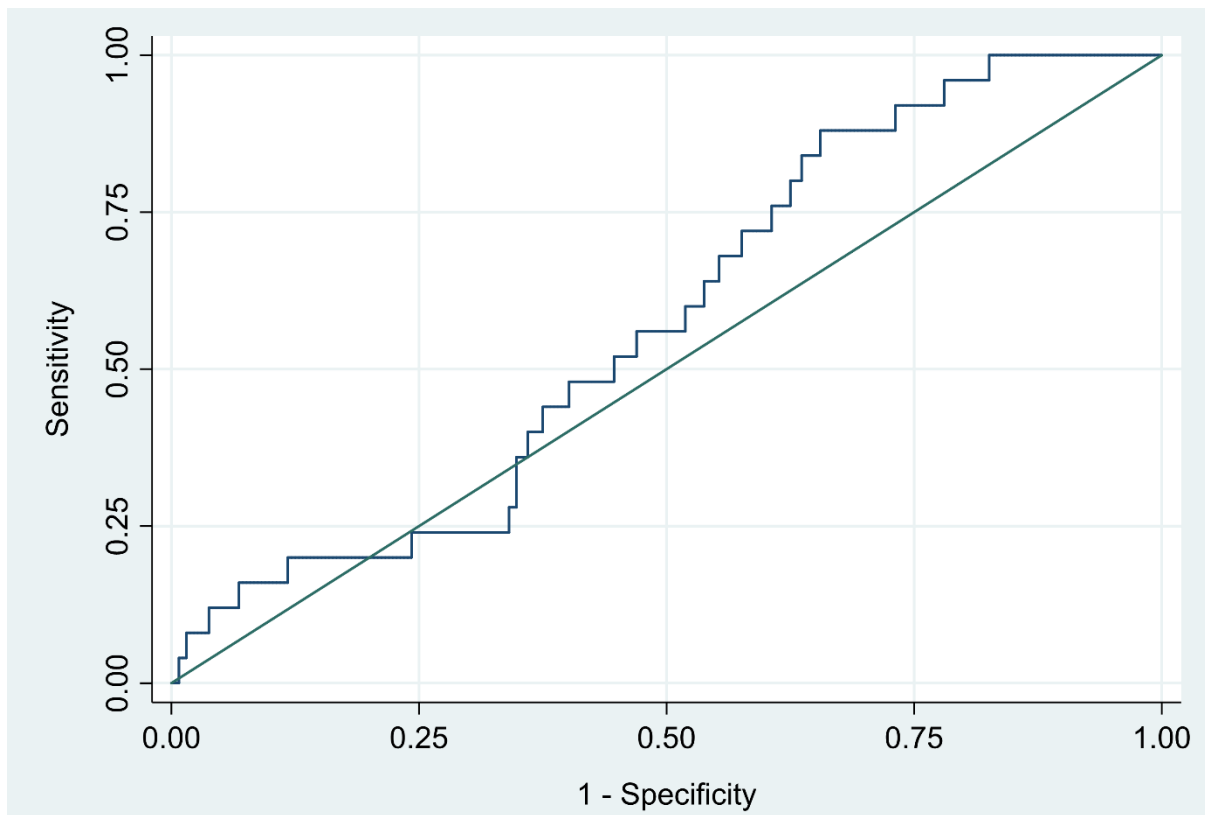
AUROC 0.58 (95% CI 0.47, 0.68).

AUROC: area under receiver operator characteristic; CI: confidence interval; PlGF: placental growth factor; sFlt-1: soluble fms-like tyrosine kinase-1

Discussion

The potential for biomarkers of placental function to help detect pregnancy complications and aid diagnosis is an area of clinical practice that is regaining momentum after a period of initial study in the 1970s and 1980s. The majority of recent work has focused on the role of PlGF or the sFlt-1/PlGF ratio in the diagnosis of preeclampsia which has led to the National Institute for Health Care and Excellence (NICE) issuing guidance on the use of these biomarkers in suspected preeclampsia.[34] As preeclampsia has its origins in placental dysfunction there is also interest in utilising placentally-derived biomarkers such as PlGF for other conditions related to

poor placental function e.g. fetal growth restriction or pregnancies that are likely to end in stillbirth[6] or fetal and neonatal compromise resulting from labour.[8]

Although there are various methods available to quantify sFlt-1 and PlGF, NICE guidance for preeclampsia specifically supports use of the Elecsys® sFlt-1 and Elecsys® PlGF assays (Roche Diagnostics; Germany).[34] Therefore, we chose to use the Elecsys® system in this work and assessed test reproducibility of both assays across six different sites. There are currently no defined or published acceptance criteria for these biomarkers based on biological variability, but we consider the performance reported here acceptable across most of the concentration range tested. Performance of the PlGF assay at concentrations <25 pg/mL was more variable and caution is advisable when interpreting results at this lower limit. This may be due to these values being at the very lowest PlGF concentrations of a very wide analytical range (3-10,000 pg/mL, limit of quantification 10 pg/mL) with a potentially greater impact of calibration curve fitting models. However, the absolute differences were small and are all below PlGF concentrations that should raise concerns in clinical practice, for example, NICE suggest a threshold of 100 pg/mL to define abnormality[34] in the context of preeclampsia. If a ratio approach is used, there is potential to increase its variability in the instances where PlGF concentrations are below 25 pg/mL and this also need to be considered in interpretation. However, when the individual sFlt-1/PlGF ratios obtained by the six sites in this study were compared to pre-eclampsia diagnosis thresholds, <2% would have changed classification. This suggested that the variable performance at low concentrations of PlGF described here would not have a major effect on clinical outcomes.

Between batch precision was acceptable for both assays with the mean CV assessed at eight sites being <5%. However, one outlier was noted at site 3 which had significantly higher

imprecision, in particular for sFlt-1, than the other sites. Most of the sites were in the method verification stage of using these biomarkers and therefore the observed precision at site 3 may have been compromised by either inconsistent handling of IQC materials or inexperience in using the assays. The most experienced laboratory (site 6) had noted that IQC performance decreased following initial opening of the bottle, even when it was kept refrigerated, and their standard practice changed to freezing aliquots of IQC material which improved observed precision. Further evaluation of sFlt-1 and PlGF stability in IQC materials is needed, particularly as the instability noted is inconsistent with the performance of these biomarkers in serum samples. This observation also highlights the need for development of independent third party IQC materials. Additionally, the clinical impact of bias and imprecision for these markers needs further consideration and this should be explored as external quality assessment schemes develop. Precision studies show that both sFlt-1 and PlGF assays can be acceptably quantitated down to the lowest concentrations tested with %CVs less than 10. We report acceptable precision and therefore a LLOQ at 12.1 pg/mL for sFlt-1 and 5.2 pg/mL for PlGF which highlights an improvement in the levels stated in the manufacturer's product sheets (15 pg/mL and 10 pg/mL respectively).

Both biomarkers also showed stability across a range of storage conditions and timeframes. This was an important consideration for samples collected in the ReMIT-2 trial since they are stored at -20^{0}C or lower for a maximum of 6 months prior to central analysis of sFlt-1 and PlGF which is being conducted as a confirmatory measure of reliability.[22] Biotin interference with the sFlt-1 assay was limited at lower concentrations, but a greater effect was seen with the PlGF assay at 60 ng/mL. Both assays showed a much larger difference with biotin at a higher concentration of 500 ng/mL. Pregnancy multivitamin tablets taken once daily contain around 50-150 µg of biotin so it is unlikely that pregnant women using such products will have blood

concentrations high enough to interfere with the sFlt-1 and PlGF assays. However, biotin is available as a single vitamin supplement at doses up to 10 mg once per day and this much higher concentration could potentially cause issues with accurate measurement of both biomarkers.

These analytical performance findings agree with previously published manufacturer's data[15] and have provided reassurance that the sFlt-1 and PlGF results generated in the ReMIT-2 trial and other multicentre clinical evaluations will be robust. This is particularly important since there is currently no External Quality Assessment scheme available in the UK for sFlt-1. It is therefore advisable to conduct similar analytical performance and stability research alongside any future trials using these biomarkers. In addition, the potential issue of biotin interference should be given due consideration in any research using the Elecsys® sFlt-1 and Elecsys® PlGF assays.

For the DTA study, although the sensitivity of the sFlt-1/PlGF ratio ≥38 to detect an adverse pregnancy outcome in women with RFM was modest and lower than that used to diagnose preeclampsia, it was comparable to, or better than, other methods currently used in clinical practice following maternal presentation with RFM e.g. ultrasound fetal biometry, liquor volume and umbilical artery Doppler >95th centile (Table 3). The combination of EFW <10th centile or sFlt-1/PlGF ratio ≥38 was included for comparison and did improve sensitivity however, as EFW <10th centile would be an indication for immediate delivery, this would not meet the inclusion criteria for the ReMIT-2 trial.[22] Using a threshold of ≥112 for the sFlt-1/PlGF ratio did not improve prognostic performance. In common with analysis of the whole FEMINA2 dataset, addition of PlGF measurement improved the sensitivity to detect adverse pregnancy outcome above use of ultrasound alone.[7] Importantly, views on this level of

sensitivity and specificity were sought from a patient and public involvement group (personal communication) and the independent Trial Steering Committee. These groups agreed that the increased sensitivity of adding sFlt-1/PlGF testing to currently available regimes without a significant reduction in specificity would aid the clinical management of women at risk of an adverse pregnancy outcome, and was therefore deemed an appropriate test to investigate further in a pilot trial. This opinion was strengthened by the fact that, in the absence of the sFlt-1/PlGF test, women with RFM would be randomised to continue with normal care in their pregnancy.[22]

The AUROC curve observed in this study of 0.58 indicates that the sFlt-1/PlGF ratio is slightly better than chance to discriminate between women with and without composite adverse pregnancy outcomes. Overall, our findings for the sFlt-1/PlGF ratio are in agreement with other analyses which suggest PlGF measurement modestly improves the detection of adverse pregnancy outcome although this may differ for specific adverse outcomes. Analysis of a cohort of 3,953 participants found a combination of EFW, PlGF and uterine artery Doppler pulsatility index had a sensitivity of 0.63 for SGA births.[35] In contrast, a combination of EFW and sFlt-1 had a sensitivity of 0.15 for emergency Caesarean section for fetal compromise during labour, and adverse outcomes after birth (low pH, Apgar score ≤7 at 5 min, NICU admission) were not predicted by any single or combination of measurements or biochemical or biophysical predictors.[35] Our recent systematic review of DTA studies found that PlGF had the best predictive accuracy for identifying a pregnancy that ends in stillbirth, but that its performance was not as good as EFW alone in prediction of a SGA infant.[9] Therefore further studies are needed to determine which biomarkers most accurately identify fetal and neonatal compromise and whether combinations of biomarkers are more successful than when used in isolation.

The current study limitations relate to sample size for both laboratory comparisons and diagnostic accuracy and the unblinded nature of the ultrasound scan findings. For the latter issue, this means that clinicians are aware of the results for EFW or abnormal liquor volume and umbilical artery Doppler which may lead to intervention thereby avoiding a potential adverse pregnancy outcome and confounding the observed diagnostic performance of these tests. Although restricted by sample size the laboratory study represents real world data from routine diagnostic laboratories and is the first data we are aware of showing how these methods perform across multiple sites in this setting. The planned sample size for the DTA study was not met meaning that the confidence intervals for the sensitivity and specificity of the sFlt-1/PlGF ratio to identify adverse pregnancy outcomes were slightly wider than planned. In addition, although modest improvements using the sF1t-1/PlGF ratio were observed in this study, the confidence intervals for the sensitivity and specificity overlapped with the other tests meaning a larger sample size would be required to identify whether the sF1t-1/PlGF ratio is better than the other tests. To address this the FEMINA3 trial is ongoing, and additional samples for DTA studies will also be obtained from the control arm of the ReMIT-2 trial.[22] Importantly, future clinical studies also need to address the impact of ethnicity, BMI and presence of co-morbidities on the test performance of the sFlt-1/PlGF ratio.

The DTA study is strengthened as samples were obtained from a single tertiary UK maternity unit serving a population with a diverse demographic and ethnic background, plus eligibility criteria were kept simple to give a degree of generalisability. In addition, the investigators determining the index test and those determining the reference standard were blinded to each other's results. Furthermore, this DTA analysis reports similar findings to a study using PlGF alone (rather than the sFlt-1/PlGF ratio) in low-risk pregnancies predicting fetal or neonatal

compromise, with AUROCs between 0.51 and 0.67 for different definitions of fetal and neonatal compromise .[36]

To the best of our knowledge, this is the first report of sFlt-1/PlGF DTA results in women with RFM in late pregnancy and is a pragmatic initial exploration of test characteristics for these assays in this population. Knowledge of test characteristics and technical variation is particularly important when applying a threshold to initiate management or treatment as small changes in absolute values can modify the ratio. The sFlt-1/PlGF threshold of ≥ 38 is being assessed further in the ongoing ReMIT-2 pilot trial to identify adverse pregnancy outcome in women with RFM, the results of which will help determine whether a much larger trial using these biomarkers in such a context is warranted.

Declaration of conflicting interests

The authors declared no potential conflict of interests with respect to the research, authorship, and/or publication of this article.

Ethical approval

Ethics approval was obtained from Oldham REC (FEMINA REC reference 08/H1011/83), Greater Manchester West REC (FEMINA2 REC reference 11/NW/0650), and Greater Manchester East REC (FEMINA3 REC reference 16/NW/0481).

Guarantor

AEPH

Contributorship

AEPH conceived the study, gained funding and ethical approval and was involved in patient recruitment. AEPH and LAB were involved in protocol development. SH and TJ were involved in method development. LB was involved in statistical analysis. LAB wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

References

1.   Hagmann H, Thadhani R, Benzing T, et al. The promise of angiogenic markers for the early diagnosis and prediction of preeclampsia. *Clinical chemistry* 2012; 58: 837-845. 2012/03/21. DOI: 10.1373/clinchem.2011.169094.
2.   Poon LC, Kametas NA, Maiz N, et al. First-trimester prediction of hypertensive disorders in pregnancy. *Hypertension (Dallas, Tex : 1979)* 2009; 53: 812-818. 2009/03/11. DOI: 10.1161/hypertensionaha.108.127977.
3.   Zeisler H, Llurba E, Chantraine F, et al. Predictive Value of the sFlt-1:PlGF Ratio in Women with Suspected Preeclampsia. *The New England journal of medicine* 2016; 374: 13-22. 2016/01/07. DOI: 10.1056/NEJMoa1414838.
4.   Cerdeira AS, Agrawal S, Staff AC, et al. Angiogenic factors: potential to change clinical practice in pre-eclampsia? *BJOG : an international journal of obstetrics and gynaecology* 2018; 125: 1389-1395. 2017/12/02. DOI: 10.1111/1471-0528.15042.
5.   MacDonald TM, Tran C, Kaitu'u-Lino TJ, et al. Assessing the sensitivity of placental growth factor and soluble fms-like tyrosine kinase 1 at 36 weeks' gestation to predict small-for-gestational-age infants or late-onset preeclampsia: a prospective nested case-control study. *BMC pregnancy and childbirth* 2018; 18: 354. 2018/09/02. DOI: 10.1186/s12884-018-1992-x.
6.   Gaccioli F, Sovio U, Cook E, et al. Screening for fetal growth restriction using ultrasound and the sFLT1/PlGF ratio in nulliparous women: a prospective cohort study. *The Lancet Child & adolescent health* 2018; 2: 569-581. 2018/08/19. DOI: 10.1016/s2352-4642(18)30129-9.
7.   Higgins LE, Myers JE, Sibley CP, et al. Antenatal placental assessment in the prediction of adverse pregnancy outcome after reduced fetal movement. *PloS one* 2018; 13: e0206533. 2018/11/06. DOI: 10.1371/journal.pone.0206533.
8.   Sherrell H, Dunn L, Clifton V, et al. Systematic review of maternal Placental Growth Factor levels in late pregnancy as a predictor of adverse intrapartum and perinatal outcomes. *European journal of obstetrics, gynecology, and reproductive biology* 2018; 225: 26-34. 2018/04/10. DOI: 10.1016/j.ejogrb.2018.03.059.
9.   Heazell AE, Hayes DJ, Whitworth M, et al. Biochemical tests of placental function versus ultrasound assessment of fetal size for stillbirth and small-for-gestational-age infants. *The Cochrane database of systematic reviews* 2019; 5: Cd012245. 2019/05/16. DOI: 10.1002/14651858.CD012245.pub2.
10.   Verlohren S, Herraiz I, Lapaire O, et al. New gestational phase-specific cutoff values for the use of the soluble fms-like tyrosine kinase-1/placental growth factor ratio as a diagnostic test for preeclampsia. *Hypertension (Dallas, Tex : 1979)* 2014; 63: 346-352. 2013/10/30. DOI: 10.1161/hypertensionaha.113.01787.
11.   Chappell LC, Duckworth S, Seed PT, et al. Diagnostic accuracy of placental growth factor in women with suspected preeclampsia: a prospective multicenter study. *Circulation* 2013; 128: 2121-2131. 2013/11/06. DOI: 10.1161/circulationaha.113.003215.
12.   Liu Y, Zhao Y, Yu A, et al. Diagnostic accuracy of the soluble Fms-like tyrosine kinase-1/placental growth factor ratio for preeclampsia: a meta-analysis based on 20 studies. *Archives of Gynecology and Obstetrics* 2015; 292: 507-518. journal article. DOI: 10.1007/s00404-015-3671-8.
13.   Verlohren S, Galindo A, Schlembach D, et al. An automated method for the determination of the sFlt-1/PlGF ratio in the assessment of preeclampsia. *American journal of obstetrics and gynecology* 2010; 202: 161.e161-161.e111. 2009/10/24. DOI: 10.1016/j.ajog.2009.09.016.

14.   Caillon H, Tardif C, Dumontet E, et al. Evaluation of sFlt-1/PlGF Ratio for Predicting and Improving Clinical Management of Pre-eclampsia: Experience in a Specialized Perinatal Care Center. *Annals of laboratory medicine* 2018; 38: 95-101. 2017/12/08. DOI: 10.3343/alm.2018.38.2.95.

15.   Schiettecatte J, Russcher H, Anckaert E, et al. Multicenter evaluation of the first automated Elecsys sFlt-1 and PlGF assays in normal pregnancies and preeclampsia. *Clinical biochemistry* 2010; 43: 768-770. 2010/03/09. DOI: 10.1016/j.clinbiochem.2010.02.010.

16.   Andersen LB, Frederiksen-Moller B, Work Havelund K, et al. Diagnosis of preeclampsia with soluble Fms-like tyrosine kinase 1/placental growth factor ratio: an inter-assay comparison. *Journal of the American Society of Hypertension : JASH* 2015; 9: 86-96. 2015/01/21. DOI: 10.1016/j.jash.2014.11.008.

17.   Nice D, Hayden K, Higgins LE, et al. Human placental lactogen and placental growth factor can differentiate small for gestational age from appropriately grown infants. *Placenta* 2014; 35: A19. DOI: https://doi.org/10.1016/j.placenta.2014.06.064.

18.   McCarthy FP, Gill C, Seed PT, et al. Performance of commercially available placental growth factor tests in women with suspected preterm pre-eclampsia; the COMPARE study. *Ultrasound in Obstetrics & Gynecology* 2018; 0. DOI: doi:10.1002/uog.19051.

19.   Trambas C, Lu Z, Yen T, et al. Characterization of the scope and magnitude of biotin interference in susceptible Roche Elecsys competitive and sandwich immunoassays. *Annals of clinical biochemistry* 2018; 55: 205-215. 2017/09/07. DOI: 10.1177/0004563217701777.

20.   Colon PJ and Greene DN. Biotin Interference in Clinical Immunoassays. *The Journal of Applied Laboratory Medicine* 2018. DOI: 10.1373/jalm.2017.024257.

21.   Grimsey P, Frey N, Bendig G, et al. Population pharmacokinetics of exogenous biotin and the relationship between biotin serum levels and in vitro immunoassay interference. *International Journal of Pharmacokinetics* 2017; 2: 247-256. DOI: 10.4155/ipk-2017-0013.

22.   Armstrong-Buisseret L, Mitchell E, Hepburn T, et al. Reduced fetal movement intervention Trial-2 (ReMIT-2): protocol for a pilot randomised controlled trial of standard care informed by the result of a placental growth factor (PlGF) blood test versus standard care alone in women presenting with reduced fetal movement at or after 36(+ 0) weeks gestation. *Trials* 2018; 19: 531. 2018/10/05. DOI: 10.1186/s13063-018-2859-1.

23.   Warrander LK, Batra G, Bernatavicius G, et al. Maternal perception of reduced fetal movements is associated with altered placental structure and function. *PloS one* 2012; 7: e34851. 2012/04/24. DOI: 10.1371/journal.pone.0034851.

24.   Heazell AE and Froen JF. Methods of fetal movement counting and the detection of fetal compromise. *Journal of obstetrics and gynaecology : the journal of the Institute of Obstetrics and Gynaecology* 2008; 28: 147-154. 2008/04/09. DOI: 10.1080/01443610801912618.

25.   Holm Tveit JV, Saastad E, Stray-Pedersen B, et al. Maternal characteristics and pregnancy outcomes in women presenting with decreased fetal movements in late pregnancy. *Acta obstetricia et gynecologica Scandinavica* 2009; 88: 1345-1351. 2009/11/03. DOI: 10.3109/00016340903348375.

26.   Norman JE, Heazell AEP, Rodriguez A, et al. Awareness of fetal movements and care package to reduce fetal mortality (AFFIRM): a stepped wedge, cluster-randomised trial. *Lancet (London, England)* 2018; 392: 1629-1638. 2018/10/03. DOI: 10.1016/s0140-6736(18)31543-5.

27.   Roche Diagnostics Limited. Advances in the management of suspected pre-eclampsia with Elecsys® sFlt-1/PlGF ratio test. 2018, https://www.healthprofessionalacademy.co.uk/docs/default-source/pre-eclampsia/pre-eclampsia-clinical-brochure.pdf?sfvrsn=22f21b22_4 Accessed 12th December 2019.

28.   Peyro Saint Paul L, Debruyne D, Bernard D, et al. Pharmacokinetics and pharmacodynamics of MD1003 (high-dose biotin) in the treatment of progressive multiple sclerosis. *Expert opinion on drug metabolism & toxicology* 2016; 12: 327-344. 2015/12/25. DOI: 10.1517/17425255.2016.1136288.

29.   Haslam S, Oakey J, Brown A, et al. A comparison of biotin interference in routine immunoassays on the Roche Cobas 8000, Beckman Coulter DXi and Siemens Advia Centaur XPT immunoassay platforms. *Clinical chemistry and laboratory medicine* 2019 2019/04/17. DOI: 10.1515/cclm-2019-0078.

30.   Dutton PJ, Warrander LK, Roberts SA, et al. Predictors of poor perinatal outcome following maternal perception of reduced fetal movements--a prospective cohort study. *PloS one* 2012; 7: e39784. 2012/07/19. DOI: 10.1371/journal.pone.0039784.

31.   Gardosi J and Francis A. Customised Centile Calculator. GROW v6.7.8.1 (UK). , www.gestation.net. Accessed 07 November 2017. (2016).

32.   Boers KE, Vijgen SMC, Bijlenga D, et al. Induction versus expectant monitoring for intrauterine growth restriction at term: randomised equivalence trial (DIGITAT). *BMJ* 2010; 341: c7087. DOI: 10.1136/bmj.c7087.

33.   Brocklehurst P and Group oboTIC. A study of an intelligent system to support decision making in the management of labour using the cardiotocograph – the INFANT study protocol. *BMC pregnancy and childbirth* 2016; 16: 10. journal article. DOI: 10.1186/s12884-015-0780-0.

34.   National Institute for Health and Care Excellence. PlGF-based testing to help diagnose suspected pre-eclampsia (Triage PlGF test, Elecsys immunoassay sFlt-1/PlGF ratio, DELFIA Xpress PlGF 1-2-3 test, and BRAHMS sFlt-1 Kryptor/BRAHMS PlGF plus Kryptor PE ratio), https://www.nice.org.uk/guidance/dg23/chapter/1-Recommendations accessed 6th December 2018] (2016).

35.   Valino N, Giunta G, Gallo DM, et al. Biophysical and biochemical markers at 35-37 weeks' gestation in the prediction of adverse perinatal outcome. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* 2016; 47: 203-209. 2015/08/01. DOI: 10.1002/uog.15663.

36.   Bligh LN, Greer RM and Kumar S. The relationship between maternal placental growth factor levels and intrapartum fetal compromise. *Placenta* 2016; 48: 63-67. 2016/11/23. DOI: 10.1016/j.placenta.2016.10.007.