Title: Ultrasound and biochemical predictors of pregnancy outcome at diagnosis of early-onset fetal growth restriction

Authors and affiliations:

Rebecca Spencer¹,², Kasia Maksym¹, Kurt Hecher³, Karel Maršál⁴, Francesc Figueras⁵, Gareth Ambler⁶, Harry Whitwell¹,³,⁸, Nuno Rocha Nenê¹, Neil J. Sebire⁹, Stefan R. Hansson⁴, Anke Diemert³,
Jana Brodzki⁴, Eduard Gratacós⁵, Yuval Ginsberg¹,¹⁰, Tal Weissbach¹,¹¹, Donald M. Peebles¹, Ian Zachary¹², Neil Marlow¹, Angela Huertas-Ceballos¹³, Anna L. David¹

¹ UCL Elizabeth Garrett Anderson Institute for Women’s Health, University College London, London, UK
² Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, UK
³ Department of Obstetrics and Fetal Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
⁴ Department of Obstetrics and Gynaecology, Institute of Clinical Sciences Lund, Skane University Hospital, Lund University, Lund, Sweden
⁵ Institut D’Investigacions Biomèdiques August Pi i Sunyer, University of Barcelona, Barcelona Center for Maternal-Fetal and Neonatal Medicine, Barcelona, Spain
⁶ Department of Statistical Science, University College London, London, UK
⁷ National Phenome Centre and Imperial Clinical Phenotyping Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK
⁸ Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK
⁹ Population, Policy & Practice Department, Great Ormond Street Institute of Child Health, University College London, London, UK

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

27/01/2023
10 Department of Obstetrics and Gynecology, Rambam Medical Centre, Haifa, Israel
11 Department of Obstetrics and Gynecology, Sheba Medical Center Tel Hashomer, Tel Aviv, Israel
12 Division of Medicine, Faculty of Medical Sciences, University College London, UK
13 Neonatal Department, University College London Hospitals NHS Foundation Trust, London, UK

Corresponding author:
Name: Rebecca Spencer
Address: LIGHT Laboratories, University of Leeds, 6 Clarendon Way, Leeds, LS2 9NL
Telephone number: 07793 355 359
Email: r.n.spencer@leeds.ac.uk

Conflicts of interest:
RS, KMak, KH, KMar, FF, GA, HW, NRN, NS, SH, AD, EG, YG, TW, IZ and ALD have no conflicts of interest to declare; NM declares consultancy fees from InfanDx over the past 12 months.

Author contributions:
RS contributed to study design, patient recruitment, data collection, sample analysis, data analysis and writing the manuscript. KMak contributed to patient recruitment and data collection. KH, KMar, FF, SH, AD, JB, EG, DMP, IZ and ALD contributed to study design, patient recruitment and data collection. GA contributed to study design and data analysis. HW contributed to sample analysis and data analysis. NRN contributed to data analysis. NS contributed to study design and sample analysis. YG and TW contributed to patient recruitment and data collection. NM and AHC contributed to study design. All authors contributed to critically revising the manuscript, give approval for publication of the work and agree to be accountable for all aspects of it.
Abstract

Background: Severe early-onset fetal growth restriction (FGR) causes significant fetal and neonatal mortality and morbidity. Predicting the outcome of affected pregnancies at the time of diagnosis is difficult, preventing accurate patient counselling. We investigated the use of maternal serum protein and ultrasound measures at diagnosis to predict fetal or neonatal death and three secondary outcomes: fetal death or delivery ≤28+0 weeks; development of abnormal umbilical artery Doppler velocimetry; slow fetal growth.

Methods: Women with singleton pregnancies (n=142, estimated fetal weights [EFWs] <3rd centile, <600g 20+0-26+6 weeks of gestation, no known chromosomal, genetic or major structural abnormalities), were recruited from four European centres. Maternal serum from the discovery set (n=63) was analysed for seven proteins linked to angiogenesis, 90 additional proteins associated with cardiovascular disease and five proteins identified through pooled liquid chromatography tandem mass spectrometry. Patient and clinician stakeholder priorities were used to select models tested in the validation set (n=60), with final models calculated from combined data.

Results: The most discriminative model for fetal or neonatal death included EFW z-score (Hadlock 3 formula/Marsal chart), gestational age and umbilical artery Doppler category (AUC 0.91, 95%CI 0.86-0.97) but was less well calibrated than the model containing only EFW z-score (Hadlock3/Marsal). The most discriminative model for fetal death or delivery ≤28+0 weeks included maternal serum placental growth factor (PIGF) concentration and umbilical artery Doppler category (AUC 0.89, 95%CI 0.83-0.94).

Conclusion: Ultrasound measurements and maternal serum PIGF concentration at diagnosis of severe early-onset FGR predict pregnancy outcomes of importance to patients and clinicians.

Trial registration: ClinicalTrials.gov NCT02097667

Funding: European Union, Rosetrees Trust, Mitchell Charitable Trust.
Introduction

The survival and growth of a fetus depends on placental provision of nutrients and waste exchange with the mother. When this system is impaired, by inadequate transformation of the uteroplacental circulation or deficits in the structure or function of the placenta, the fetus fails to reach their growth potential (1, 2). The resulting fetal growth restriction (FGR) may be diagnosed antenatally on the basis of an ultrasound determined low estimated fetal weight (EFW) for gestational age; either <3rd centile or <10th centile with abnormal Doppler ultrasound indices in the uterine (UtA) and/or umbilical (UmA) arteries (3) (4, 5). Early-onset FGR, occurring before 32 weeks of gestation, carries significant risks of stillbirth, neonatal morbidity and mortality, neurodevelopmental impairment, and long-term health problems (6-13). There is currently no treatment that can improve fetal growth in utero; instead, management involves monitoring the pregnancy and timing delivery to balance the risks of stillbirth and prematurity (4, 14-16).

An important question when developing novel therapies for early-onset FGR is which pregnancies to include in early-phase clinical trials. There is a balance to be struck between identifying pregnancies that are sufficiently severely affected to justify the possible risks of the intervention but not so severely affected that there is no potential to determine efficacy. Numerous studies have investigated predictive markers for the development of FGR (17-22), but far fewer have studied the prediction of pregnancy outcome when early-onset FGR is diagnosed. Lack of knowledge about pregnancy outcome in FGR makes it difficult to optimise the inclusion criteria for clinical trials, but it also leaves pregnant patients and their partners with a considerable burden of uncertainty (23, 24).

The EVERREST Project aims to carry out a phase I/IIa trial of maternal vascular endothelial growth factor (VEGF) gene therapy for early-onset FGR (25). The greatest potential for benefit is in pregnancies at the threshold of viability, for whom our current management option, preterm...
delivery, is not possible or is very high risk. In preparation for a clinical trial of a novel therapeutic,
we established a multicentre prospective study to characterise the natural history of early-onset
FGR, choosing an extreme phenotype in which the estimated fetal weight (EFW) was <3rd centile and
<600g between 20+0 and 26+6 weeks of gestation (henceforth referred to as ‘severe’ early-onset
FGR) (26).

The aim of this work was to prospectively identify and validate ultrasound and serum biochemical
factors that could be used to predict fetal or neonatal death in pregnancies affected by severe early-onset FGR. These could subsequently be used to select the most appropriate women for a first-in-human study of a novel therapeutic to treat FGR, and to better counsel women and their partners about pregnancy outcome. To this end, we asked patients and clinicians to assess the value of our primary and secondary outcomes, based on which we then selected models for validation.

Unsupervised parenclitic network analysis by pregnancy outcome and functional network analysis of proteins associated with pregnancy outcomes were also performed to maximise the utility of the proteomic data with the aim of providing insights into the underlying pathophysiology.

**Results**

The discovery set, recruited between March 2014 and September 2016, comprised 63 pregnant
participants (Figure 1 & Table 1; supplementary data Table 1). Follow-up for the ascertainment of
study outcomes was completed in November 2016. The validation set, recruited between October 2016 and January 2020, comprised 60 pregnant participants, with follow-up for the ascertainment of
study outcomes completed in March 2020. There were no significant differences in maternal
demographics, pregnancies characteristics or pregnancy outcomes between the discovery and
validation sets (Table 1; supplementary data Table 2 & Figure 1). Overall, 42 (34%) of the
pregnancies ended in the primary outcome of fetal or neonatal death (within the first 28 days of
life). For the three secondary outcomes, only fetal death or delivery ≤28+0 weeks of gestation could
Page 5

27/01/2023
be ascertained for all pregnancies, occurring in 58 cases (47%). The UmA Doppler velocimetry was
normal (<95th centile for gestation) at enrolment in 46 participants, of whom 21 (46%)
subsequently developed abnormal UmA Doppler measurements. Fetal growth trajectory (based on
change in percentage weight deviation over a period or two weeks or more) could be assessed
for 104 pregnancies (85%), with the remaining pregnancies ending in fetal death or delivery before a
2-week interval was reached. Forty-one of these 104 fetuses (39%) demonstrated slow fetal growth
(worsening of weight deviation of ≥10 percentage points). A smaller proportion of fetuses
demonstrated slow fetal growth in the validation set (31%) than in the discovery set (47%).

Figure 1: Flow diagram of participant eligibility and enrolment across the four EVERREST Prospective
Study centres from 10th March 2014 to 30th January 2020 for the discovery and validation sets.
Table 1: Characteristics and outcomes of the discovery, validation and combined participant sets.

Discovery and validation sets were compared using 2-sided t tests for symmetrical continuous variables, Mann-Whitney U tests for skewed continuous variables and chi square or Fisher’s exact tests (where specified) for categorical variables.

<table>
<thead>
<tr>
<th>Maternal characteristics</th>
<th>Discovery n=63</th>
<th>Validation n=60</th>
<th>p value</th>
<th>Combined n=123</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age, mean years (SD)</td>
<td>33.7 (6.0)</td>
<td>33.3 (6.7)</td>
<td>0.72</td>
<td>33.5 (6.3)</td>
</tr>
<tr>
<td>Primiparous, n (%)</td>
<td>43 (68)</td>
<td>37 (62)</td>
<td>0.44</td>
<td>80 (65)</td>
</tr>
<tr>
<td>BMI, median (IQR)</td>
<td>24.9 (22.7-28.6)</td>
<td>26.4 (22.8-31.0)</td>
<td>0.36</td>
<td>25.7 (22.8-30.0)</td>
</tr>
<tr>
<td>Ethnicity, n (%)1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>42 (67)</td>
<td>27 (46)</td>
<td>0.16</td>
<td>69 (57)</td>
</tr>
<tr>
<td>Black</td>
<td>10 (16)</td>
<td>15 (25)</td>
<td></td>
<td>25 (20)</td>
</tr>
<tr>
<td>Asian</td>
<td>10 (16)</td>
<td>15 (25)</td>
<td></td>
<td>25 (20)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (2)</td>
<td>2 (3)</td>
<td></td>
<td>3 (2)</td>
</tr>
<tr>
<td>Essential hypertension, n (%)</td>
<td>10 (16)</td>
<td>6 (10)</td>
<td>0.33</td>
<td>16 (13)</td>
</tr>
<tr>
<td>Pre-eclampsia at enrolment, n (%)2</td>
<td>4 (6)</td>
<td>5 (8)</td>
<td>0.47</td>
<td>9 (7)</td>
</tr>
</tbody>
</table>

Enrolment ultrasound measurements

<table>
<thead>
<tr>
<th>Maternal characteristics</th>
<th>Discovery n=63</th>
<th>Validation n=60</th>
<th>p value</th>
<th>Combined n=123</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gestational age at enrolment, median weeks+days (IQR) [range]</td>
<td>23+6 (22+3-25+1)</td>
<td>23+5 (22+3-24+4)</td>
<td>0.48</td>
<td>23+5 (22+3-24+5)</td>
</tr>
<tr>
<td>EFW at enrolment, median grams (IQR) (29)</td>
<td>392 (281-503)</td>
<td>387 (280-448)</td>
<td>0.61</td>
<td>389 (281-484)</td>
</tr>
<tr>
<td>EFW z-score at enrolment, median (IQR) (29, 30)</td>
<td>-3.0 (-3.6 to -2.5)</td>
<td>-3.2 (-3.8 to -2.7)</td>
<td>0.17</td>
<td>-3.1 (-3.7 to -2.6)</td>
</tr>
<tr>
<td>Mean UTA PI &gt;95th centile at enrolment, n (%)2 (27)</td>
<td>49 (79)</td>
<td>49 (82)</td>
<td>0.71</td>
<td>98 (80)</td>
</tr>
<tr>
<td>UmA PI &gt;95th centile at enrolment, n (%)2 (27)</td>
<td>32 (51)</td>
<td>36 (60)</td>
<td>0.26</td>
<td>78 (55)</td>
</tr>
<tr>
<td>Absent or reversed UmA end-diastolic flow at enrolment, n (%)3 (27)</td>
<td>18 (29)</td>
<td>25 (42)</td>
<td>0.11</td>
<td>43 (35)</td>
</tr>
<tr>
<td>MCA PI &lt;5th centile at enrolment, n (%)4 (27)</td>
<td>10 (16)</td>
<td>7 (12)</td>
<td>0.41</td>
<td>17 (15)</td>
</tr>
<tr>
<td>DV a wave absent or reversed at enrolment, n (%)4</td>
<td>5 (8)</td>
<td>4 (7)</td>
<td>0.51</td>
<td>9 (8)</td>
</tr>
</tbody>
</table>

Pregnancy outcomes

<table>
<thead>
<tr>
<th>Maternal characteristics</th>
<th>Discovery n=63</th>
<th>Validation n=60</th>
<th>p value</th>
<th>Combined n=123</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-eclampsia at any point in pregnancy, n (%)</td>
<td>24 (39)</td>
<td>17 (33)</td>
<td>0.51</td>
<td>41 (36)</td>
</tr>
<tr>
<td>Gestational age at diagnosis of stillbirth or delivery of live birth, median weeks+days (IQR) [range]</td>
<td>28+2 (26+3-34+0)</td>
<td>28+2 (26+3-33+2)</td>
<td>0.90</td>
<td>28+2 (26+3-33+2)</td>
</tr>
<tr>
<td>Female fetus/infant, n (%)7</td>
<td>36 (57)</td>
<td>24 (43)</td>
<td>0.19</td>
<td>60 (50)</td>
</tr>
<tr>
<td>Caesarean delivery, n (%)</td>
<td>42 (67)</td>
<td>37 (63)</td>
<td>0.65</td>
<td>79 (65)</td>
</tr>
<tr>
<td>Live births, n (%)</td>
<td>47 (75)</td>
<td>43 (72)</td>
<td>0.71</td>
<td>90 (73)</td>
</tr>
<tr>
<td>Live births ≤28+0 weeks, n (%)</td>
<td>15 (32)</td>
<td>14 (33)</td>
<td>0.95</td>
<td>29 (32)</td>
</tr>
<tr>
<td>Live births &gt;37 weeks, n (%)</td>
<td>8 (17)</td>
<td>10 (23)</td>
<td>0.46</td>
<td>18 (20)</td>
</tr>
<tr>
<td>Caesarean delivery for live births, n (%)</td>
<td>42 (89)</td>
<td>37 (88)</td>
<td>0.85</td>
<td>79 (89)</td>
</tr>
<tr>
<td>Neonatal deaths, n (%)8</td>
<td>5 (11)</td>
<td>4 (9)</td>
<td>0.56</td>
<td>9 (10)</td>
</tr>
</tbody>
</table>
Birth weight z-score for live births, mean (SD) (30)

<table>
<thead>
<tr>
<th>Study outcomes</th>
<th>n=63</th>
<th>n=60</th>
<th>n=123</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fetal or neonatal death, n (%)</td>
<td>21 (33)</td>
<td>21 (35)</td>
<td>0.85</td>
</tr>
<tr>
<td>Death or delivery ≤28+0 weeks of gestation, n (%)</td>
<td>30 (48)</td>
<td>28 (47)</td>
<td>0.92</td>
</tr>
<tr>
<td>Slow fetal growth, n (%)</td>
<td>26 (47)</td>
<td>15 (31)</td>
<td>0.083</td>
</tr>
<tr>
<td>Development of UmA PI &gt;95th centile, n (%)</td>
<td>12 (46)</td>
<td>9 (45)</td>
<td>0.94</td>
</tr>
</tbody>
</table>

1n=1 missing from validation, 2n=1 missing from discovery, 3n=1 missing from validation, 4n=5 missing from discovery and n=1 missing from validation, 5n=3 missing from discovery and n=1 missing from validation, 6n=1 missing from discovery and n=8 missing from validation, 7n=5 missing from validation, 8Fisher’s exact test. DV=ductus venosus, EFW=estimated fetal weight, MCA=middle cerebral artery, PI=pulsatility index, UmA=umbilical artery, UtA=uterine artery.

Ultrasound measurements as predictors of fetal or neonatal death and death or delivery ≤28+0 weeks of gestation in the discovery set: The best ultrasound predictor of fetal or neonatal death was EFW z-score, either as calculated using the Hadlock 3 formula and Marsal chart (EFW<sub>HLM</sub>: AUC 0.81, 95% CI 0.69-0.93) or the Intergrowth formula and chart (EFW<sub>Int</sub>: AUC 0.83, 95% CI 0.71-0.95). UmA category (<95th centile; >95th centile with positive EDF; absent EDF; reversed EDF. AUC 0.75, 95% CI 0.62-0.88) and slow fetal growth (AUC 0.70, 95% CI 0.56-0.83) were also fair predictors. UmA category was the best predictor of death or delivery ≤28+0 weeks (AUC 0.80, 95% CI 0.70-0.91), with mean UtA PI (AUC 0.77, 95% CI 0.65-0.89) and Intergrowth EFW z-score (AUC 0.73, 95% CI 0.60-0.85) also fair predictors (supplementary data Tables 3 & 4).

Proteomics: Mass spectrometry (MS) profiling of pooled samples gave quantitative information for 200 protein groups (sets of proteins that cannot be distinguished based on peptide sequences), from which human placental lactogen (HPL), fibronectin, pregnancy-specific beta-1 glycoprotein (PSG1), serum amyloid A (SAA) and leucyl-cystinyl aminopeptidase (LNPEP) were selected for individual validation, based on the scoring system outlined in the methods.
Univariate associations between maternal serum protein concentrations and outcomes in the discovery set: Four proteins were undetectable in most of the samples: VEGF-A, natriuretic peptides B (BNP), melusin, poly[ADP-ribose] polymerase 1 (PARP1). These were excluded from further prediction analyses. The associations between the remaining 98 proteins and the four pregnancy outcomes are summarised in Figure 2. Placental growth factor (PlGF) and HPL concentration were significantly associated with fetal or neonatal death (after Benjamini-Hochberg correction), with fold-changes of 0.52 in pregnancies ending in fetal or neonatal death compared with pregnancies ending in neonatal survival. The concentrations of nine proteins were significantly associated with death or delivery ≤28+0 weeks (after correction). The greatest magnitudes of fold-changes were seen for PlGF (0.28), HPL (0.45), and PSG1 (0.48).
Figure 2: Volcano plots showing the statistical significance and magnitude of the associations between the 98 proteins and four pregnancy outcomes in the discovery set. Associations tested with 2-sided t tests for symmetrical data and Mann-Whitney U tests for skewed data. Dotted line indicates a p value of 0.05. Dashed line indicates the Benjamini-Hochberg cut-off with a 5% false discovery rate (p=0.0015 for fetal or neonatal death, p=0.005 for death or delivery ≤28+0 weeks). See supplementary data Table 5 for full protein names.
Parenclitic network analysis of the discovery set: Both the networks for fetal or neonatal death and death or delivery ≤28+0 weeks contained clusters centred around HPL (clusters 6 and 4, Figure 3).

These clusters also contained pentraxin-related protein PTX3, spondin-2 (SPON2) and thrombomodulin (TM) and contained or were linked to decorin (DCN). The network for death or delivery ≤28+0 weeks also contained a cluster centred around PlGF (cluster 2). For all three of the networks that included fetal sex, similar clusters emerged that contained renin (REN), angiopoietin 1 (ANG1), dickkopf-related protein 1 (DKK1) and platelet-derived growth factor subunit beta (PDGFβ), contained or were linked to pregnancy-associated plasma protein A (PAPP-A) and in two of the three networks included lectin-like oxidized LDL receptor 1 (LOX1). Networks and associated dendrograms for the development of abnormal umbilical artery Dopplers and slow fetal growth are provided in supplementary data Figures 4 & 5.
Figure 3: Parenclitic networks for clusters generated based on: (A) fetal or neonatal death (B) death or delivery ≤28+0 weeks. See supplementary data Table 5 for full protein names and supplementary data Figures 2 & 3 for associated dendrograms.
Model selection by stakeholders: None of the single variable or multivariable models for predicting slow fetal growth performed well enough to warrant validation (AUCs <0.70). For the three remaining outcomes, an online survey was performed to ascertain the priorities of clinicians and patients in predicting outcomes. Forty-five clinicians from 18 countries (of 173 contacted, 26%) and seven patients from the UK who had experienced a pregnancy complicated by severe early-onset FGR (of 36 contacted, 19%) responded (supplementary data Table 6). The prediction of fetal or neonatal death and death or delivery ≤28+0 weeks were considered important or very important by all patients and were also rated highly by clinicians for the purposes of patient counselling and clinical management (Figure 4). Patients and clinicians marginally prioritised sensitivity over specificity for most outcomes (supplementary data Figure 6). For the prediction of the development of abnormal UmA PI, patients universally prioritised sensitivity while clinicians marginally prioritised specificity for patient counselling.

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Very important</th>
<th>Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient priorities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient counselling</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: The proportion of patients who ranked our three pregnancy outcomes as ‘important’ or ‘very important’ and the proportion of clinicians who ranked them as ‘important’ or ‘very important’ for clinical management and for patient counselling. UmA PI=umbilical artery pulsatility index.
Based on the survey results, the model performance metrics and the assay reliability, models containing the variables listed in Tables 2 and 3 were selected for validation. For the prediction of death or delivery ≤28+0 weeks, models including HPL marginally outperformed models including PIGF. However, possibly related to the short processing time, the commercial HPL ELISA had high intra-assay variability in our hands (mean coefficient of variation 8.0%, SD 7.3%, 28% requiring repeat analysis for coefficient of variation >10%). Because of this, and the existence of clinically approved tests for PIGF, models including PIGF were selected for validation.

**Model validation:** Five of the seven protein models (Table 2) and all five of the models containing ultrasound measurements (Table 3), generated in the discovery set, were successfully validated, with AUCs included in the AUC 95% CIs generated from the discovery cross-validation estimates.

**Addition of pregnancy characteristics:** Validated models were not significantly improved by the addition of maternal BMI, maternal age, maternal ethnicity or fetal sex. Adding gestational age at enrolment significantly improved the models containing EFW_{HM} alone (LR test p=0.0001) and EFW_{HM} with UmA category (LR test p<0.00005) to predict fetal or neonatal death (supplementary data Table 7). The addition of ‘pre-eclampsia at enrolment’ significantly improved all validated models predicting fetal death or delivery ≤28 weeks of gestation (supplementary data Table 7).

**PIGF values for maximum likelihood ratios:** Receiver operating characteristic (ROC) curves are shown in Figures 5 and 6. Model constants and coefficients, along with optimal cut points for positive and negative likelihood ratios and correct classification are provided in supplementary data Tables 8 and 9. Serum PIGF concentration <14.2 pg/ml predicted fetal or neonatal death with a positive likelihood ratio of 18.3, sensitivity of 45% and specificity of 98% and correctly classified 80% of participants. Serum PIGF concentration <14.5 pg/ml predicted death or delivery ≤28+0 weeks with a positive likelihood ratio of 24.7, sensitivity of 38% and specificity of 98% and correctly classified 70% of participants.
Table 2: Model validation for predicting adverse pregnancy outcomes: the seven pre-specified models containing maternal serum proteins. The model not validated is shaded. Validated models include estimates from the combined discovery and validation sets.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Variable(s)</th>
<th>Discovery (with LOOCV)</th>
<th>Validation</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AUC</td>
<td>95% CI</td>
<td>AUC</td>
</tr>
<tr>
<td>Fetal or neonatal death</td>
<td>PIGF</td>
<td>0.75</td>
<td>0.62-0.88</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>PIGF &amp; lymphotactin</td>
<td>0.84</td>
<td>0.73-0.95</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>PIGF, lymphotactin &amp; fibronectin</td>
<td>0.85</td>
<td>0.74-0.96</td>
<td>0.69</td>
</tr>
<tr>
<td>Death or delivery ≤28+0 weeks</td>
<td>PIGF</td>
<td>0.86</td>
<td>0.76-0.96</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>PIGF &amp; pre-eclampsia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PIGF &amp; PSG1*</td>
<td>0.91</td>
<td>0.82-0.99</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>PIGF, PSG1 &amp; pre-eclampsia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development of abnormal UmA PI</td>
<td>PIGF</td>
<td>0.84</td>
<td>0.50-0.95</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>PIGF &amp; fibronectin</td>
<td>0.88</td>
<td>0.74-1.00</td>
<td>0.65</td>
</tr>
</tbody>
</table>

*Discovery AUC 0.906, validation 95% CI 0.688-0.912. LOOCV=leave-one-out cross-validation, PI=pulsatility index, PIGF=placental growth factor, PSG1=pregnancy-specific beta-1 glycoprotein, UmA=umbilical artery. Models were generated and tested using the natural log of PIGF in pg/ml and centered and scaled values for lymphotactin normalised protein expression (NPX) on a log2 scale.
Table 3: Model validation for predicting adverse pregnancy outcomes: models containing ultrasound measurements, maternal serum protein concentrations and pregnancy characteristics, and their final AUCs from the combined discovery and validation sets.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Variable(s)</th>
<th>Discovery (with LOOCV)</th>
<th>Validation</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AUC</td>
<td>95% CI</td>
<td>AUC</td>
</tr>
<tr>
<td>Fetal or neonatal death</td>
<td>EFW&lt;sub&gt;HM&lt;/sub&gt; z-score</td>
<td>0.78</td>
<td>0.66-0.91</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>EFW&lt;sub&gt;HM&lt;/sub&gt; z-score &amp; GA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EFW&lt;sub&gt;int&lt;/sub&gt; z-score</td>
<td>0.80</td>
<td>0.67-0.93</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>EFW&lt;sub&gt;HM&lt;/sub&gt; z-score &amp; UmA category&lt;sup&gt;1&lt;/sup&gt;</td>
<td>0.78</td>
<td>0.64-0.91</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>EFW&lt;sub&gt;HM&lt;/sub&gt; z-score, UmA category &amp; GA&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death or delivery ≤28+0 weeks</td>
<td>UmA category&lt;sup&gt;1&lt;/sup&gt;</td>
<td>0.78</td>
<td>0.67-0.89</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>UmA category &amp; pre-eclampsia&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UmA category &amp; PlGF&lt;sup&gt;1&lt;/sup&gt;</td>
<td>0.89</td>
<td>0.81-0.97</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>UmA category, PlGF &amp; pre-eclampsia&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>1</sup>n=1 missing from validation set. EFW<sub>HM</sub>=estimated fetal weight calculated using Hadlock 3 formula (29) with z-score calculated using Marsal reference chart (30), EFW<sub>int</sub>=estimated fetal weight and z-score calculated using Intergrowth formulae (31), GA=gestational age at enrolment in days, LOOCV=leave-one-out cross-validation, PlGF=placental growth factor, UmA=umbilical artery.
Figure 5: Comparison of the receiver operating characteristic (ROC) curves for the models predicting fetal or neonatal death. EFW-HM=estimated fetal weight calculated using Hadlock 3 formula (29) with z-score calculated using Marsal reference chart (30), EFW-Intergrowth=estimated fetal weight and z-score calculated using Intergrowth formula and reference chart (31), GA=gestational age at enrolment, PlGF=placental growth factor concentration, UmA=umbilical artery Doppler category (0=pulsatility index <95th centile, 1=pulsatility index >95th centile, 2=absent end-diastolic flow, 3=reversed end-diastolic flow).

Figure 6: Comparison of the receiver operating characteristic (ROC) curves for the models predicting fetal death or delivery ≤28+0 weeks of gestation. PET=pre-eclampsia at enrolment, PlGF=placental growth factor concentration, PSG1=pregnancy-specific glycoprotein 1 normalised protein expression, UmA=umbilical artery Doppler category (0=pulsatility index <95th centile, 1=pulsatility index >95th centile, 2=absent end-diastolic flow, 3=reversed end-diastolic flow).
Alternative EFW formulae: Although the EFW z-score calculated using the Intergrowth formula and chart gave the highest AUC for predicting fetal or neonatal death, the Intergrowth formula for estimating fetal weight performed poorly in our sample, especially at lower fetal weights. For the 21 livebirths with a birthweight <600g and an EFW performed within seven days of delivery, the Intergrowth formula overestimated birthweight by a mean of 47% (SD 14%), in contrast to the Hadlock 3 formula which overestimated birthweight by a mean of 25% (SD 10%, supplementary data Figure 8). For all 67 livebirths with an EFW performed within seven days of delivery, the Intergrowth formula overestimated birthweight by a mean of 29% (SD 20%) and the Hadlock 3 formula overestimated birthweight by a mean of 15% (SD 13%). As might be expected, using the EFW calculated from one formula in the model derived from the other had a substantial negative impact on calibration (supplementary data Figure 9).

Re-analysis of the combined sets: Combining the centred and scaled data from the discovery and validation sets, the strongest associations with both fetal or neonatal death, and death or delivery ≤28+0 weeks were the previously observed negative associations with PlGF (p=1.4x10^{-8} and p=3.0x10^{-11}) and HPL (p=1.3x10^{-7} and p=1.4x10^{-10}) (Figure 7). The evidence for the negative association between PSG1 and both outcomes was strengthened, as was the evidence for negative associations between matrix metalloproteinase-12 (MMP12) and programmed cell death 1 ligand 2 (PDL2) and death or delivery ≤28+0 weeks. None of the proteins showed an association with the development of abnormal UmA Dopplers or slow fetal growth at a Benjamini-Hochberg 5% false discovery rate (supplementary data Figure 10).
Figure 7: Volcano plots showing the statistical significance and magnitude of associations between fetal and neonatal death and death or delivery ≤28+0 weeks and the centred and scaled concentrations of the 93 proteins from the discovery and validation sets combined. Associations tested with 2-sided t tests. Dotted line indicates p=0.05, short-dashed line indicates Benjamini-Hochberg cut-off with a 5% false discovery rate (A p=0.0048, B p=0.012), long-dashed line indicates Benjamini-Hochberg cut-off with a 1% false discovery rate (A p=0.00032, B p=0.0013). See supplementary data Table 10 for full protein names and individual -log10 p values.

Functional analysis of the proteins associated with fetal or neonatal death at a 5% false discovery rate demonstrated co-expression of HPL and GH. Expanding the network to include intervening proteins resulted in clusters sharing GO biological processes of growth hormone receptor signalling, VEGF signalling and calcitonin family receptor signalling, with proteins in the latter two clusters also involved in angiogenesis and regulation of angiogenesis (Figure 8). Proteins associated with death or
delivery ≤28+0 weeks showed multiple interactions, predominantly centred on fibronectin. Shared GO biological processes included those relating to growth (regulation of angiogenesis, cellular response to growth factors and growth hormone receptor signalling pathway via jak-stat), immune function (leucocyte migration, inflammatory response, and positive regulation of T cell activation) or both (regulation of cell adhesion, positive regulation of nik/NF-kappaB signalling) (Figure 9).

Figure 8: An expanded functional network demonstrating interactions and shared GO biological processes of the proteins associated with fetal or neonatal death in the combined data set at a Benjamini-Hochberg false discovery rate of 5%. Red=growth hormone receptor signalling pathway, yellow=vascular endothelial growth factor signalling pathway, green=calcitonin family receptor signalling pathway, light blue=regulation of angiogenesis, dark blue=angiogenesis. See supplementary data Table 5 for full protein names. Analysis, graphic and legend from STRING (Swiss Institute of Bioinformatics)(32).
Figure 9: Functional interactions and shared GO biological processes of the proteins associated with death or delivery ≤28+0 weeks in the combined data set at a Benjamini-Hochberg false discovery rate of 5%. Red = growth hormone receptor signalling pathway via jak-stat, orange = inflammatory response, yellow = leukocyte migration, light green = extracellular matrix, dark green = regulation of cell adhesion, light blue = regulation of angiogenesis, dark blue = cellular response to growth factor stimulus, pink = positive regulation of nik/NF-kappaB signalling, purple = positive regulation of T cell activation. See supplementary data Table 5 for full protein names. Analysis, graphic and legend from STRING (Swiss Institute of Bioinformatics)(32).

The three best performing LOOCV models using the combined centred and scaled data all included pro-adrenomedullin (ADM) for predicting fetal or neonatal death, PIGF and HPL for predicting death or delivery ≤28+0 weeks and Platelet-derived growth factor subunit B (PDGFB) for predicting the development of abnormal UmA Doppler measurements (supplementary data Table 12). The emergence of ADM in the models predicting fetal or neonatal death was consistent with the significant association present in the combined but not the discovery sets. In contrast, PDGFB did not emerge in the prediction of abnormal UmA Doppler.
not show significant univariate associations with the development of abnormal UmA Dopplers in the discovery, validation, or combined data sets.

**Predicting gestational age of livebirth or diagnosis of fetal death and interval from enrolment to livebirth or diagnosis of fetal death:** Twelve protein and ultrasound measurements showed an association with the gestational age at which the pregnancies ended in livebirth or fetal death, at a 1% Benjamini-Hochberg false discovery rate (supplementary data Table 13). The best model to predict gestational age at livebirth or fetal death included PIGF and sflt1 concentrations, MMP12 and IL1RL2 NPX and UmA category at enrolment (Figure 10). Eight protein and ultrasound measurements showed an association with the interval between enrolment and either livebirth or the diagnosis of fetal death at a 1% Benjamini-Hochberg false discovery rate (supplementary data Table 13). The best model to predict the interval between enrolment and livebirth or fetal death included PIGF and sflt1 concentrations, MMP12 and decorin NPX, UmA category and gestational age at enrolment (Figure 10). Both models accounted for 68% of the variation in the outcomes they were predicting but had 95% prediction intervals of 40 days, limiting their clinical utility. Sparser models, including PIGF and sflt1 concentrations and UmA category to predict gestational age at livebirth or fetal death, and these same variables plus gestational age at enrolment to predict interval to livebirth or fetal death, had only slightly wider 95% prediction intervals of 42 days (supplementary data Figure 11).
Figure 10: (A) Predicted versus actual gestational age of either livebirth or diagnosis of fetal death, based on the model containing PlGF and sflt1 concentrations, decorin and matrix metalloproteinase 12 normalised protein expression and umbilical artery Doppler category. (B) Predicted versus actual interval from enrolment to either livebirth or diagnosis of fetal death, based on the model containing PlGF and sflt1 concentrations, decorin and matrix metalloproteinase 12 normalised protein expression, umbilical artery Doppler category and gestational age at enrolment. Green filled circles=pregnancies ending in livebirth, red hollow circles=pregnancies ending in fetal death, dotted lines=95% prediction intervals.

Placental histological classification: Placental samples for histological examination were available for 55 pregnancies (45%); these had similar characteristics and outcomes to the pregnancies without available samples (supplementary data Tables 14 & 15). The only statistically significant difference was a higher proportion of female fetuses among the pregnancies that had placental samples than those that did not (63% vs 41%, p=0.016). Forty-five (82%) placentas showed evidence of placental pathology, with 39 (71%) classified as maternal vascular malperfusion (MVM), three (5%) as villitis of unknown aetiology (VUE), one (2%) as fetal vascular malperfusion (FVM) and two (4%) as non-specific dysmorphic villi. Twelve of the 14 placental samples from pregnancies ending in stillbirth showed MVM, while the three available samples from pregnancies ending in neonatal death showed
VUE, FVM and dysmorphic villi. Mean UtA PI, maternal serum PlGF and maternal serum PAPP-A at enrolment all differed significantly between pregnancies with subsequent MVM and pregnancies without MVM. In contrast, none of the umbilical artery parameters studied (UmA category at enrolment; the occurrence of UmA PI >95th centile at any point before delivery; the occurrence of absent or reversed UmA end-diastolic flow at any point before delivery) showed evidence of an association with placental histological classification of MVM (supplementary data Table 16).

Discussion

Principle findings and significance

To our knowledge this is the first study to use a discovery science approach, combining ultrasound and biochemical parameters, to identify and validate prognostic markers at the time of diagnosis of severe early-onset FGR. These findings can be used to inform personalised counselling and management of affected pregnancies with outcomes of importance to patients and clinicians. Furthermore, by providing alternative thresholds that prioritise positive and negative likelihood ratios and maximum correct categorisation, eligibility criteria for clinical trials of novel therapeutics can be adapted depending on the perceived risk: benefit ratio of the intervention.

Our secondary analyses, including parenclitic network analysis, functional enrichment analysis and triangulation with placental histological classification, provide a deeper characterisation of this unique case series. Some of these findings support and enhance our existing understanding of placental FGR, such as the interplay between angiogenesis, immune cells, and the extracellular matrix (33-35). Other findings offer new avenues for investigation, such as the parenclitic network cluster around fetal sex, which includes proteins related to pericyte function (36).

Findings in the context of existing literature

Given that ultrasound assessment of biometry and Doppler velocimetry forms the mainstay of identifying and monitoring FGR, it is unsurprising that EFW z-score and umbilical artery category
were validated as predictors of fetal or neonatal death and fetal death or delivery ≤28+0 weeks respectively (37-39). A secondary analysis of 105 pregnancies from the UK placebo-controlled trial of sildenafil citrate for early-onset FGR (STRIDER; EFW or AC <10th centile with absent or reversed UmA end-diastolic flow at 22+0-29+6 weeks) identified EFW as an independent predictor of livebirth (OR per 100g 4.3, 95%CI 2.3-8.0, p<0.001) and overall survival (OR per 100g 2.9, 95%CI 1.8-4.4, p<0.001) (40). What is less expected is that absent or reversed ductus venosus a wave was a poor predictor of fetal or neonatal death in our participants (AUC 0.59, 95% CI 0.53-0.66, see supplementary material Table 4), in contrast to the results of previous studies (41, 42). This may reflect a change in clinical practice since these studies were published. Their findings led to ductus venosus waveform becoming an important factor in timing of delivery in extremely pre-term FGR (14, 43), which may have altered the natural history of the disease by prompting delivery before stillbirth could occur.

A limitation of using ultrasound parameters is their potential for variation. In the case of Doppler velocimetry this includes interobserver variability, temporal variation due to factors such as maternal and fetal movement, and variation in umbilical artery waveforms between arteries and along the length of the cord (44-47). There is also considerable variation between different Doppler reference ranges, both in terms of the values of their ‘normal’ ranges and their methodological quality (48). In the case of EFW z-score, variation arises from interobserver variability in measuring biometry, variation in formulae used to generate the EFW and variation in charts used to determine the z-score for gestational age (49-51). Despite the methodological limitations of the Hadlock 3 formula, a recent study of 65 pregnancies with early-onset FGR delivered within seven days of ultrasound assessment found it gave a better combination of systematic and random error than 20 other formulae tested (52).

Several recent studies have highlighted the potential utility of PlGF concentration to predict outcomes in SGA and FGR pregnancies. In a case series of 173 singleton pregnancies with a
customised EFW <10th centile between 20+0 and 31+6 weeks, the sflt1/PIGF ratio at diagnosis was an excellent predictor of delivery <30 weeks (AUC 0.96, 95% CI 0.93-0.99) and <34 weeks (AUC 0.94, 95% CI 0.89-0.98) and a good predictor of a composite adverse perinatal outcome (AUC 0.83, 95% CI 0.77-0.90) (53). Similarly, in 116 singleton pregnancies with early-onset FGR (customised EFW <3rd centile or customised EFW <10th centile with abnormal UmA and/or UtA Doppler velocimetry; <32+0 weeks) and positive UmA end-diastolic flow ending in livebirth, women with an sflt1/PIGF ratio ≥85 were significantly more likely to deliver within one, two, three and four weeks from ratio measurement than women with an sflt1/PIGF ratio <85 (54). Composite neonatal morbidity and neonatal admission were also significantly higher following pregnancies with a sflt1/PIGF ratio ≥85 (53.8% vs 28.6% p=0.04; 97.5% vs 67.9% p<0.01). More strikingly, in a series of 130 singleton pregnancies with SGA (AC or EFW <10th centile), fetal demise only occurred in pregnancies with a PIGF <10th centile for gestational age at any time between 16 and 36 weeks (12/65 vs 0/65, p<0.0001) (55).

While these studies revealed the PIGF results to the managing clinicians, similar results have been found in studies where PIGF was not revealed. The secondary analysis of STRIDER UK trial participants, mentioned above, found significant associations between pregnancy outcomes and both the sflt1/PIGF ratio and PIGF alone (40). Higher PIGF concentrations and a lower sflt1/PIGF ratios were associated with greater overall survival (PIGF coefficient 3.67, p<0.001; ratio coefficient 0.51, p=0.002) and later gestation at birth (PIGF coefficient 1.4, p<0.001; ratio coefficient -0.99, p<0.001). Similarly, in a multinational case series of 411 pregnancies, PIGF <5th centile at the time of suspected FGR (AC <10th centile from 20+0 weeks) had an 87.5% sensitivity and 62.8% specificity for predicting stillbirth (56). PIGF <12 pg/ml was associated with a shorter interval to delivery than PIGF >5th centile (13.0 vs 29.5 days, p<0.0001).

Our finding that placental histological classification of maternal vascular malperfusion was significantly associated with lower maternal PIGF concentration and higher mean uterine artery PI at
diagnosis of early-onset FGR, but not with UmA Doppler measurements, was in keeping with the results of previous studies (56-59). Agrawal et al. have recently reported that MVM, unlike other placental pathologies, is characterised by raised mean uterine artery PI and a gradual decline in PlGF as the pregnancy progresses (57). Triunfo et al. found in SGA pregnancies (EFW <10th) identified between 30 and 34 weeks of gestation, a pattern of placental histopathology they termed ‘placental underperfusion’, was most strongly associated with lower PlGF, measured at the time of diagnosis (58). Benton et al. also found low PlGF to be a better predictor of placental pathology than umbilical artery resistance index or abdominal circumference centile (56).

Strengths and limitations

The strengths of this multicentre study are that it was carried out prospectively in academic health science centres with fetal medicine experts trained on a common ultrasound protocol, and level 3 perinatal care available for delivery. Participants and their fetuses/neonates were extensively phenotyped at study entry, for the duration of the pregnancy and postnatally, and we report temporally validated results. All pregnancies were managed according to local guidelines, although these were broadly consistent, in line with national and international guidelines (4, 16, 60, 61) and current RCT evidence (e.g. the TRUFFLE trial). This introduces variation, but potentially better reflects real-world practice and hence adds external validity. All serum analysis was carried out after pregnancy outcomes were obtained, using a proteomic discovery science approach which did not assume associations with outcome, but also included additional analysis of proteins anticipated to be related to pregnancy outcome in placental insufficiency. Placental histological classification was blinded to pregnancy outcomes and included control and non-FGR preterm placental samples, to remove some potential bias. Finally, our inclusion of stakeholders to guide model selection means that their predictive value is most important to patients and clinicians.
Our relatively narrow inclusion criteria are both a strength and a limitation, in that they have allowed us to focus on a specific clinical group but have limited our sample size and the generalisability of our findings. The sample size means our study was underpowered to demonstrate small or medium effects and our estimates have wider confidence intervals than larger studies (62). The exclusion of pregnancies <3rd centile but >600g limits the number of pregnancies from 24+6 weeks of gestation to which our findings can be applied and the exclusion of pregnancies with known genetic, chromosomal, and structural differences means our findings cannot be applied to the whole spectrum of FGR. Generalisability is also limited to healthcare settings with comparable neonatal care provision and outcomes, given their impact on neonatal survival and decision making for iatrogenic preterm delivery. Clinicians managing the pregnancies were not blinded to ultrasound measurements, and indeed many management decisions will have been influenced by ultrasound findings. This could have biased the apparent associations between ultrasound variables and pregnancy outcomes, either artificially strengthening or weakening them.

**Future directions**

Ideally, our findings should be independently and externally validated. Given the incidence of FGR ≤28+0 weeks this would require another multicentre study. Further research is also needed to determine whether the use of these models would have benefit in practice, both on the psychological wellbeing of parents and on the use of health resources. Future studies to identify and validate predictive models in early-onset SGA (EFW <10th centile <32+0 weeks of gestation) would allow application to a wider population. This would complement the work currently being done in the PLANES study, which is investigating the impact of revealed PlGF in SGA from 32+0 weeks (63). Finally, our primary outcome of fetal or neonatal death provides only short-term information, and data collection for 2-year neurodevelopmental outcomes is on-going.
Conclusion

In conclusion, our study provides validated models for predicting fetal or neonatal death and fetal death or delivery ≤28+0 weeks of gestation based on ultrasound and maternal serum protein measurements at the time of diagnosis of severe early-onset FGR. The EFW z-score and umbilical artery Doppler velocimetry were the best performing ultrasound parameters, but are vulnerable to inter-rater variability, variation in formulae and reference ranges and temporal variation. The biomarker PlGF was the best performing maternal serum protein for predicting both pregnancy outcomes and maternal vascular malperfusion. This identification of a specific pathological phenotype may be useful for targeting future potential therapies.

Methods

This study is reported according to the ‘strengthening the reporting of observational studies in epidemiology’ (STROBE) guidelines (64) for cohort studies and the ‘transparent reporting of a multivariable prediction model for individual prognosis or diagnosis’ (TRIPOD) guidelines (65).

Study design and setting: The EVEREST Prospective Study was a multicentre prospective cohort study recruiting pregnant women from four European tertiary referral centres: University College
Study population: Full details of the protocol have been published previously (26). In brief, pregnant women were eligible if they had a singleton fetus with an ultrasound estimated fetal weight (EFW) <600g and <3rd centile according to local criteria between 20+0 and 26+6 weeks of gestation.

Exclusion criteria were a known abnormal karyotype or major fetal structural abnormality at enrolment (66), indication for immediate delivery, preterm rupture of membranes before enrolment, maternal HIV or hepatitis B or C infection, maternal age under 18 years, any medical or psychiatric condition which compromised a woman’s ability to participate and lack of capacity to consent. Pregnancies with a known congenital infection were not recruited and for the purposes of this analysis pregnancies ending in termination were excluded.

Outcomes: The primary outcome was fetal or neonatal death (≤28 days of life). Secondary outcomes were: fetal death or delivery ≤28+0 weeks of gestation; slow fetal growth, defined as a worsening of weight deviation of ≥10 percentage points over a two-week interval (including before and after enrolment) or equivalent trajectory over a longer period (28); and the development of abnormal UmA Dopplers, defined as development of UmA PI >95th centile in pregnancies where UmA PI was ≤95th centile at enrolment (27). Ascertainment for outcomes of this study was possible, at the latest, by 29 days of life. Follow-up for neonatal morbidity and infant health and neurodevelopment to the age of 2 years continues.

All pregnancies were managed according to the local fetal medicine unit protocols. Pre-eclampsia was defined according to International Society for the Study of Hypertension in Pregnancy (ISSHP) criteria (67), meaning that, given the presence of FGR, any woman developing new onset hypertension after 20+0 weeks of gestation was classified as having pre-eclampsia rather than pregnancy-induced hypertension. Formalin-fixed placental samples were classified according to Amsterdam consensus criteria by a single assessor (NS) (68). To minimise bias, study placental
samples were mixed with placental samples from healthy term pregnancies and pregnancies delivering spontaneously preterm, with NS blinded to pregnancy phenotype and outcome during the assessment.

**Ultrasound measurements:** All ultrasound examinations were performed by staff trained and validated to the common EVERREST prospective study protocol (26). At each ultrasound scan Doppler velocimetry of the umbilical artery (UmA), uterine artery (Uta), middle cerebral artery (MCA), ductus venosus (DV) and umbilical vein was performed (69). Local EFW formulae and centile charts were used to determine study eligibility, but for consistency all EFWs were recalculated using the Hadlock 3 formula (incorporating head circumference, abdominal circumference and femur length), with z-scores recalculated using the Marsal chart for descriptive data (supplementary data Equations 1-3) (29, 30). EFWs and z-scores were also recalculated using Intergrowth formulae for analysis (supplementary data Equations 4 & 5) (31). The effect of alternative Doppler reference charts was explored, with similar results to those presented (70-74).

**Blinding:** Maternal serum protein concentrations were not available to clinicians, participants or researchers during the pregnancy, as all samples were analysed after complete primary outcome data had been ascertained. Serum PlGF and sflt-1 concentrations were not used as part of clinical care in any of the study centres during the period of recruitment.

**Sample collection:** Maternal blood was collected at study enrolment in BD Vacutainer® serum separating tubes and processed according to the manufacturer’s instructions. 500μl serum aliquots were frozen and stored at -80°C. Placental samples, for Amsterdam criteria categorisation, were collected from two areas of each placenta, midway between the cord insertion and margin in areas free from macroscopic infarcts or lesions. Samples were rinsed in PBS, formalin-fixed, wax-embedded, sectioned and stained with H&E.

**Measurement of a priori candidate biomarkers in maternal serum:** PlGF and sflt-1 concentrations were measured using Elecsys® electrochemiluminescence immunoassays on a Cobas® e411 analyser
(Roche Diagnostics). The normalised protein expression (NPX) of 90 additional proteins associated with cardiovascular disease was measured using the Olink® Cardiovascular II proximity extension assay (full list of proteins in supplementary data Table 17). In the discovery set but not the validation set VEGFA, VEGFD, VEGFR2, neuropilin 1 and endoglin were measured in triplicate using Quantikine® colorimetric sandwich ELISAs (R&D Systems).

**Identification of novel candidate biomarkers in maternal serum using liquid chromatography and tandem mass spectrometry:** Five pooled serum samples were created on the following basis: (1) pregnancies ending in fetal or neonatal death (2) pregnancies ending in neonatal survival with delivery <37+0 weeks of gestation (3) pregnancies ending in neonatal survival with delivery 37+0 weeks of gestation or more (4) slow fetal growth trajectory (5) normal fetal growth trajectory.

Pooled serum samples were depleted of 12 high-abundance proteins using Proteome Purify™ 12 resin, as per the manufacturer’s instructions (R&D Systems), concentrated using Vivaspin® 500 5kDa Molecular Weight Cut-Off columns (GE Healthcare), reduced with 10mM tris(2-carboxyethyl)phosphine hydrochloride then alkylated with 7.5mM iodoacetamide. Pooled samples were digested using a trypsin/Lys-C mix, labelled with Tandem Mass Tags™ (Thermo Fisher Scientific) and combined(75). The combined sample underwent two-dimensional high-performance reverse-phase liquid chromatography and tandem mass spectrometry. In the first dimension, samples were fractionated into 30 at high pH using a Poroshell 300 Extend C18 column (Agilent), following which fractions 1-4 were combined with fractions 27-30 respectively due to low abundance in the first four fractions. The second fractionation was performed on the Ultimate 3000 nano-liquid chromatography system using Acclaim™ PepMap™ 100 C18 pre-columns and Acclaim™ PepMap™ 100 C18 Nano-LC columns run in tandem with analysis on the LTQ (linear trap quadrupole) Orbitrap XL™ 2.5.5 (all Thermo Fisher Scientific). A blank calibration sample was run after every three fractions, and a standard sample of known mass run after every six fractions, for quality control.
Proteins were identified using Proteome Discover V1.4 software (ThermoFisher Scientific) to search the human Swiss-Prot database with the Mascot search engine (Matrix Science Ltd.). Proteins were scored on variability, peptide count, ubiquity, ratio between pools and consistent trend across pools (supplementary data Tables 18 & 19). Expression pattern clusters, based on standardised and raw quantification ratios, were generated using the Graphical Proteomics Data Explorer (GPRoX) platform. Based on their scores and expression clusters, five candidate proteins were selected and measured in individual samples using ELISAs. Fibronectin, PSG1 (both R&D Systems) and HPL (DRG International) were measured in the discovery and validation sets while SAA (R&D Systems) and LNPEP (Cloud-clone) were measured in the discovery set only.

**Priority survey and model selection:** An online survey was sent to patients and clinicians asking their opinion on the importance of different pregnancy outcomes and, for each outcome, whether they would prioritise sensitivity or specificity. Models were selected on the basis of the survey results and the model performance metrics described below. Protein models were published online prior to the validation data analysis.

**Sample size:** Since this work involved discovery of novel biomarkers, a formal *a priori* sample size calculation was not possible. Before analysing the discovery set it was determined that this sample of n=63 with 21 fetal or neonatal deaths gave an 80% power to detect a standardised effect size of 0.9 (large) to a significance level of 0.05 (76).

**Statistics:** Data analysis was performed using STATA/MP 16.1 software (StataCorp LLC, College Station, USA) unless otherwise specified. Descriptive and investigative variables were tested for skew and kurtosis (77, 78) and handled as symmetrical if there was no evidence of either. PI GF, sflt1, endoglin, VEGFD, NP1, HPL, SAA and LNPEP were transformed to their natural logs and multiplex data was analysed as provided, on a log2 scale. Characteristics of the discovery and validation sets were compared using chi-square tests (categorical data), Fisher’s exact tests (binary data with sparse
outcomes), 2-sided t tests (symmetrical continuous data) and Mann-Whitney U tests (skewed continuous data).

Missing data for BMI (n=5) were imputed using chain equations. Umbilical artery (UmA) PI at enrolment was systematically missing (n=16), with most missing cases having absent or reversed end-diastolic flow (EDF, n=15). Umbilical artery Doppler velocimetry was therefore handled as an interval variable, ‘UmA PI category’, , where 0=UmA PI ≤95th centile, 1=UmA >95th centile with positive EDF, 2=absent EDF, 3=reversed EDF. Where uterine artery (UtA) PI at enrolment was missing (n=10) a mean UtA PI below or above the 95th centile could be inferred in nine cases where the mean UtA PI was consistently normal (n=1) or abnormal (n=8) respectively at scans prior to and after enrolment and UtA PI values were imputed using multiple imputation. Associations between ultrasound measurements and both fetal or neonatal death and death or delivery ≤28+0 weeks of gestation, were analysed using logistic regression. Univariate associations between protein concentrations or NPX and outcomes were assessed using 2-sided t tests, Mann-Whitney U tests and logistic regression, with Benjamini-Hochberg procedures to account for multiple comparisons.

**Model development:** Two-protein models for the development of abnormal UmA Dopplers and two- and three-protein models for the other three pregnancy outcomes, with internal validation using leave-one-out cross-validation (LOOCV), were compared based on AUC, specificity for 90% sensitivity, sensitivity for 90% specificity, F1 score, Matthews correlation coefficient (MCC) and precision-recall characteristics (PRROC) AUC. ROC curves were generated with the pROC R package (version 1.18.0, [https://cran.r-project.org/web/packages/pROC/index.html](https://cran.r-project.org/web/packages/pROC/index.html)). 95% confidence intervals for AUCs were determined by stratified bootstrapping. PRROC curves were generated with the MLmetrics R package (version 1.1.1, [https://cran.r-project.org/web/packages/MLmetrics/index.html](https://cran.r-project.org/web/packages/MLmetrics/index.html)). Models with variance inflation factors of five or more were excluded using the following R package: [https://cran.r-project.org/web/packages/car/index.html](https://cran.r-project.org/web/packages/car/index.html), version 3.1-0. Two-variable models predicting fetal or neonatal death and death or delivery ≤28+0 weeks of gestation containing ultrasound parameters with or without PlGF or HPL (as the proteins showing the strongest
associations with these outcomes) were compared in the same way. Outcomes and protein models to be validated were published on the study registry prior to analysis of the validation data.

**Parenclitic network analysis:** Parenclitic networks of the 102 proteins were generated for each of the four pregnancy outcomes. For each outcome, two-dimensional kernel density estimations were generated for every pair-combination of variables in ‘controls’ (pregnancies without the outcome).

Individual networks were then generated for each ‘case’, with linkages created if a pair-wise relationship of variables differed from the control distribution by more than a given threshold (79).

These individual case networks were then combined. VEGF-A, BNP, PARP1 and melusin were included as binary variables of ‘detectable’ or ‘not detectable’. Booking body mass index (BMI) and fetal sex were included as variables in the networks, except for ‘development of abnormal UmA Dopplers’, where the sample was not large enough to accommodate them.

**Model validation:** Concentrations of HPL and PSG1, as measured by ELISA, and NPX values for the Olink multiplex proteins showed substantial variation in centrality and spread between the discovery and validation sets. To account for this, values of each protein were centred to a mean of 0 and scaled to an SD of 1 in the discovery set and validation set separately. These centred and scaled values were used for subsequent analyses, including model validation. Concentrations of PIGF, sflt1 and fibronectin did not require transformation.

Models generated from the discovery set were run on data from the validation set and were considered validated if the 95% CI for the validation estimate of AUC included the LOOCV AUC estimate from the discovery set. For validated models, data from both sets were combined to give final test characteristics. Likelihood ratio tests were used to determine whether the addition of pregnancy characteristics (maternal BMI, maternal age, maternal ethnicity, fetal sex, gestational age at enrolment and pre-eclampsia at enrolment) significantly improved the validated models. Model calibration was assessed by plotting predicted probability against observed frequency of outcome.
**Functional interactions:** Centred and scaled data from the discovery and validation sets were combined to retest univariate associations with the primary and secondary outcomes. Proteins showing a significant association at a 5% Benjamini-Hochberg false discovery rate were explored for physical and functional interactions and for enrichment of GO biological processes, relative to the background of all proteins measured, using STRING (Swiss Institute of Bioinformatics)(32). Where no enrichment was detected shared GO biological processes were identified through comparison to the whole genome.

**Modelling pregnancy duration:** Protein and ultrasound measurements from the combined discovery and validation sets were tested for their association with gestational age at livebirth or diagnosis of fetal death and interval from enrolment to livebirth or diagnosis of fetal death using linear regression. Variables showing a significant association at a 1% Benjamini-Hochberg false discovery rate were used to create linear models predicting these outcomes in a stepwise fashion. Maternal age, BMI, ethnicity, gestational age at enrolment, pre-eclampsia at enrolment and fetal sex were tested for model improvement. Model fit was tested by assessing variance inflation factors for multicollinearity, assessing the distribution of the residuals for heteroscedasticity and outliers, and looking for observations with high leverage. UmA and UtA Doppler velocimetry, PIGF concentration, HPL concentration and PAPP-A NPX were tested for their associations with placental histological classification using logistic regression.

**Study approval:** Ethical approval was provided by the National Research Ethics Service Committee London - Stanmore in the UK (REC reference: 13/LO/1254), the Hospital Clinic of Barcelona’s Clinical Research Ethics Committee in Spain (Reg: HCB/2014/0091), the Regional Ethical Review Board in Lund for Sweden, (DNr 2014/147) and the Ethics Committee of Hamburg Board of Physicians in Germany (PV4809). This study was conducted according to the Declaration of Helsinki principles and written informed consent was given by all participants before enrolment.
Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 305823, the Rosetrees Trust and the Mitchell Charitable Trust in memory of Shoshana Mitchell Glynn. This research has been supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre (RS, NM, ALD). NRN would like to thank the support from Cancer Research UK (C12077/A26223).

This work would not have been possible without the contribution of the late Professor John Timms, Elizabeth Garrett Anderson Institute for Women’s Health, University College London. His contribution to the design, conduct and analysis of the study was invaluable, and we would have been honoured to have him as a co-author. We would also like to thank all of the women and families who took part in this study as well as staff at the UCL Comprehensive Clinical Trials Unit (Anna Morka, Jade Dyer, Helen Knowles, Steve Hibbert, Kate Maclagan), Gina Buquis, Jade Okell, Dr Mark Lees, Dr Carlo Rossi, Dr Tara Krishnan, Dr Roberta Morris, Dr Sarah Guillon, Richard Gunu, Dr Eva Sedlak, Professor Alexei Zaikin and Dr Oleg Blyuss.

Data access statement: The full data set will not be made publicly available because the degree of detailed phenotyping could allow individual patient identification. Limited data sharing may be possible, with the agreement of the consortium, on request to RS or ALD.

References


Page 38

27/01/2023


