Sovio, U., Goulding, N., McBride, N., Cook, E., Gaccioli, F., Charnock-Jones, D. S., Lawlor, D. A., & Smith, G. C. S. (2020). A maternal serum metabolite ratio predicts fetal growth restriction at term. *Nature Medicine*, *26*, 348-353. https://doi.org/10.1038/s41591-020-0804-9

A maternal serum metabolite ratio predicts fetal growth restriction at term

Ulla Sovio, PhD[a,b*]

Neil Goulding, PhD[c,d,e]

Nancy McBride, MSc[c,d,e]

Emma Cook, BSc[a]

Francesca Gaccioli, PhD[a,b]

D Stephen Charnock-Jones, PhD[a,b]

Debbie A Lawlor, PhD[c,d,e]

Gordon C S Smith, DSc[a,b*]

[a]Department of Obstetrics and Gynaecology, University of Cambridge; NIHR Cambridge Biomedical Research Centre, Cambridge, United Kingdom

[b]Centre for Trophoblast Research (CTR), Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom

[c]NIHR Bristol Biomedical Research Centre, Bristol, United Kingdom

[d]MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, United Kingdom

[e]Population Health Sciences, Bristol Medical School, Bristol, United Kingdom

[*]**Corresponding authors:**

Dr Ulla Sovio, Department of Obstetrics and Gynaecology, University of Cambridge, Cambridge, UK.

Professor Gordon C S Smith, Department of Obstetrics and Gynaecology, University of Cambridge, Cambridge, UK.

Fetal growth restriction (FGR) is the major single cause of stillbirth[1] and is also associated with neonatal morbidity and mortality[2,3], impaired health and educational achievement in childhood[4,5] and with a range of diseases in later life[6]. Effective screening and intervention for FGR is an unmet clinical need. Here, we performed UPLC-MS/MS metabolomics on maternal serum at 12, 20, and 28 weeks of gestational age (wkGA) using 175 cases of term FGR and 299 controls from the POP study, conducted in Cambridge, UK, to identify predictive metabolites. Internal validation using 36 wkGA samples demonstrated that a ratio of the products of the relative concentrations of two positively associated metabolites (1-(1-enyl-stearoyl)-2-oleoyl-GPC and 1,5-anhydroglucitol) to the product of the relative concentrations of two negatively associated metabolites (5α-androstan-3α,17α-diol disulfate and N1,N12-diacetylspermine) predicted FGR at term. The ratio had approximately double the discrimination as compared to a previously developed angiogenic biomarker[7], the sFLT1:PlGF ratio (AUC 0.78 versus 0.64, *P*=0.0001). We validated the predictive performance of the metabolite ratio in two sub-samples of a demographically dissimilar cohort, Born in Bradford, conducted in Bradford, UK (*P*=0.0002). Screening and intervention using this metabolite ratio in conjunction with ultrasonic imaging at around 36 wkGA could plausibly prevent adverse events through enhanced fetal monitoring and targeted induction of labor.

A large proportion of adverse events associated with FGR are unrelated to maternal risk factors[8] and this has motivated research on screening for FGR. However, given that the primary intervention for FGR is early delivery, screening and intervention could cause harm by iatrogenic prematurity in false positives[9]. This may explain why the most promising approach to screening for FGR, namely, universal ultrasound, does not result in better outcomes[10]. Consequently, the primary method of screening for FGR in low risk women in the USA, UK and many other countries remains clinical examination, such as measurement of the symphyseal-fundal height[11]. We have previously argued that one approach to the current impasse is to focus initial efforts on screening and intervention at term[12]. One third of all stillbirths occur at term and infants with a birth weight <3rd percentile at term

have an eight fold increased risk of antepartum stillbirth[3]. Moreover, early term delivery, while not completely benign[13], has less potential for harm than preterm delivery[4]. However, ultrasound is less effective as a screening test for term FGR than preterm FGR[7,14].

We performed untargeted metabolomics in maternal serum to identify metabolites for use in screening tests for term FGR using 175 cases and a random sample of 299 women from the Pregnancy Outcome Prediction study (**Supplementary Table 1 and Extended Data Fig. 1**). Term FGR was defined as birth weight <3rd percentile or birth weight 3rd to <10th percentile combined with the lowest decile of fetal abdominal growth velocity[15,16]. The case-cohort study design employed combines the advantages of a cohort study with the efficiency of a case control study[17]. Longitudinal mixed effects regression of the log transformed multiple of the median (MoM) for each metabolite (see Methods) in the 12, 20 and 28wkGA samples generated a composite *P* value at 20/28wkGA for each metabolite and the *P* value distribution was skewed towards lower values (Kolmogorov-Smirnov test *P*=0.002, **Extended Data Fig. 2**). The 100 metabolites with the lowest *P* values were evaluated further. Internal validation was accepted if the *P* value at 36wkGA was below the Bonferroni-corrected *P* value of 5×10^{-4}: 22 were validated using this highly conservative threshold (**Supplementary Table 2**). A correlation matrix (**Supplementary Table 3**) demonstrated that the levels of some of these metabolites were correlated with each other, with maternal characteristics and with the sFLT1:PlGF ratio, an angiogenic biomarker ratio that we have previously shown is predictive for FGR[7]. As the aim of the study was to generate novel predictors, we included the sFLT1:PlGF ratio, maternal age, body mass index (BMI) as well as the 22 validated metabolites in a forward stepwise logistic regression. Nine of the 22 metabolites were independently predictive of FGR and five of these improved the prediction over the sFLT1:PlGF ratio based on area under the ROC curve, estimated using 1000-fold bootstrapping to account for over-fitting. One of the five, the tobacco metabolite cotinine N-oxide, was excluded from further analyses since smoking is well recognized to be associated with FGR and can be assessed by other means.

The associations between the four remaining metabolites at 36wkGA and term FGR are shown in **Table 1**. The two metabolites which were positively associated with the risk of term FGR (1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1) and 1,5-anhydroglucitol) had a declining trend throughout pregnancy (**Figs. 1a and 1b**). By contrast, the two metabolites which were negatively associated with term FGR (5α-androstan-3α,17α-diol disulfate and N1,N12-diacetylspermine) had an increasing trend throughout pregnancy (**Figs. 1c and 1d**). Hence, each of the four associations observed with FGR represented attenuation of the physiological change observed in normal pregnancy.

In order to assess the predictive ability of measurements of the four metabolites in combination, we calculated a ratio of the product of the MoMs of the two positively associated metabolites over the product of the MoMs of the two negatively associated metabolites (**Table 1**). We considered the possibility that this approach, while appealing in its simplicity, might reduce the information which could potentially be obtained by including all four measurements in a multivariable statistical model. However, the AUC for the ratio was similar to the AUC for the predicted probability derived from a multivariable logistic regression model fitted to the four metabolites (0.778 vs 0.783, respectively, *P*=0.59), hence all further analysis employed the metabolite ratio. The AUC for the metabolite ratio for term FGR at 36wkGA was 0.78 compared with 0.64 for the sFLT1:PlGF ratio (**Fig. 2a**).

For external validation, we first analysed 970 fasting plasma samples (20 FGR and 950 controls) obtained between 24 and 28wkGA from the Born in Bradford (BiB) cohort where there was information on the birthweight percentile (**Extended Data Fig. 3**). The BiB study did not include a 36wkGA blood test and the 24-28wkGA sample was the latest available. The two primary exposures pre-defined in the analysis plan were the metabolite ratio and 5α-androstan-3α,17α-diol disulfate as a sole predictor and we employed a Bonferroni corrected *P* value threshold of 0.025 (one-sided). Given the BiB sample size, we had an 87% and 89% statistical power, respectively, to identify similar

associations between the two primary exposures and birth weight percentile <3rd as were observed in the POP study (using the 28wkGA sample and the same definition of FGR). When we applied the same definition of FGR to both studies, the AUC for the metabolite ratio was similar when measured at 24-28wkGA from fasting maternal plasma in the BiB study (0.68, 95% CI: 0.55 to 0.81, *P*=0.0029 [one-sided], **Fig. 2b**) as it was when measured at 28wkGA in non-fasting maternal serum from the POP study (0.72, 95% CI: 0.67 to 0.77, **Fig. 2b**). Subsequently, a second sample became available from the BiB study, consisting of 41 cases of FGR and a comparison group of 1513 (**Extended Data Fig. 4**). The AUC for the metabolite ratio in the second sample was 0.62 (95% CI 0.54 to 0.71, *P*=0.0018 [one-sided], **Fig. 2b**). Logistic regression, as per the analysis plan, yielded *P* values (one-sided) of 0.01 in the first sample, 0.004 in the second sample and 0.0002 when both samples were pooled. Thus, despite the facts that the BiB cohort sub-samples were demographically highly dissimilar to the POP study cohort (**Supplementary Tables 4 and 5**), that the measurement was made earlier in pregnancy, and that the sample obtained was different (fasting plasma rather than non-fasting serum), the metabolite ratio was validated in two separate sub-samples from the BiB cohort. Moreover, the strength of the association with FGR was similar to the POP study samples obtained at comparable gestational windows (**Fig. 2b**).

We next assessed the diagnostic effectiveness of the metabolite ratio, estimated fetal weight (EFW) and the sFLT1:PlGF ratio measured at 36wkGA, in relation to FGR at term in the POP study (**Table 2**). EFW and sFLT1:PlGF were classified by previously defined thresholds (<10th percentile and >38, respectively[11,18]) and the metabolite ratio was classified as >85th percentile, as this is the equivalent of sFLT1:PlGF >38 at 36wkGA[7]. The combination of EFW and the metabolite ratio had the highest diagnostic odds ratio. We explored a range of cut-points for EFW and metabolite ratio (**Extended Data Fig. 5**). The combination of EFW<20th and metabolite ratio >80th percentile identified ~56% of the term FGR cases while giving a false positive rate of ~5%. The positive likelihood ratio (LR+) was ~11 and a third of women who tested positive experienced the outcome. Defining women as screen

positive on the basis of one or both of the predictors being present (i.e. either EFW<20th and/or metabolite ratio >80th percentile) was, as expected, less predictive of FGR (positive LR 2.5, [2.2 to 2.9]) and PPV 10.3% [8.2%-12.8%]). However, the absence of both (i.e. EFW≥20th and metabolite ratio ≤80th percentile) was highly effective in ruling out the disease with an extremely low negative LR (0.07 [0.03-0.15]) and extremely high NPV (99.7% [99.3%-99.9%]). Hence, using the lowest 20% of EFW and highest 20% of metabolite ratio, the population could be divided into three groups: one (containing 7.3% of women) with a very high risk of FGR (33.1%, both tests positive), one (containing 33.3% of women) with an intermediate risk of FGR (5.2%, only one test positive), and one (containing 59.4% of women) with a very low risk of FGR (0.3%, neither test positive).

When analysed by phenotype of FGR, the metabolite ratio was more strongly predictive than the sFLT1:PlGF ratio of FGR without preeclampsia. Conversely, the sFLT1:PlGF ratio was more strongly predictive than the metabolite ratio of FGR with preeclampsia (**Fig. 2c**). The superior performance of the metabolite ratio over the sFLT1:PlGF ratio was also observed when FGR was defined on the basis of birth weight percentile combined with the presence of ultrasonic features of FGR (**Supplementary Table 6**). The metabolite ratio was similarly predictive when women with different characteristics were compared (**Fig. 2d**). In the BiB sample 1, the one-sided *P* value for the other main exposure (5α-androstan-3α,17α-diol disulfate) was 0.03, i.e. just above the pre-defined, Bonferroni corrected threshold of 0.025. However, in the BiB sample 2, the one-sided *P* value was 0.001. Moreover, the analysis plan pre-specified eight secondary exposures and these were also not validated based on the Bonferroni corrected threshold of 0.00625 in either BiB sample (**Supplementary Tables 7 and 8** and **Extended Data Fig. 6**). However, in most cases the direction and magnitude of the association was similar to the associations observed in the POP study.

The four metabolites identified are all plausible markers of FGR. However, only one of these, N1,N12-diacetylspermine, had previously been described as predictive of FGR and arose from a prior

analysis of the current dataset, where the focus was on feto-placental sex and the mother's serum metabolome[16]. Low levels of 1,5-anhydroglucitol have previously been associated with increased birth weight in women with diabetes mellitus[19]. Therefore, it is plausible that higher levels are associated with FGR, although understanding the mechanistic basis of this observation will require further studies. 5α-androstan-3α,17α-diol disulfate is a steroid metabolite which was first identified in the faeces of pregnant women[20]. In our analyses, it was strongly correlated with estriol-3-sulfate which is itself highly correlated with estriol in pregnant women[21]. This is relevant as, prior to the widespread implementation of ultrasound, low levels of estriol were used to identify placental insufficiency[22,23]. Finally, 1-(1-enyl-stearoyl)-2-oleoyl-GPC is a plasmalogen and elevated placental levels of a plasmalogen derivative have previously been described in preeclampsia[24] but there are no previous studies of FGR, to our knowledge. Interestingly, for all four metabolites, the direction of the association with FGR was the opposite of the direction of the association with advancing gestational age, i.e. if the metabolite increased with advancing pregnancy, low levels were associated with FGR and *vice versa*. This observation suggests that FGR is associated with attenuation of the physiological metabolic changes associated with normal pregnancy, and this could reflect placental growth or trophoblast function. We speculate that the most likely explanation is dysregulation of normal placental metabolism, as placental dysfunction is thought to underlie many cases of FGR[25].

In the current study, we used a statistically rigorous approach to biomarker discovery employing both internal and external validation. The *P* value for the ratio at 36wkGA was $1.1x20^{-21}$ hence there is very strong evidence supporting the association between the metabolite ratio and term FGR within the POP study cohort. The evidence supporting the association was further strengthened by external validation in the BiB cohort. The samples of women comprising the BiB cohort were ~53% Pakistani ethnicity, ~37% lived in an area in the lowest quintile of socioeconomic status, the majority of women had previous births and all women were screened for gestational diabetes using a 75g fasting oral glucose tolerance test. The high level of dissimilarity between the two cohorts makes it

very unlikely that the association is due to some unmeasured confounder. The simplest explanation for validation in BiB is that the association between the metabolite ratio and FGR is true and it reflects underlying biological processes which are common to all humans. We speculate that successful identification of this generalizable association in the POP study reflects the strict statistical methods used for the selection of variables combined with the simple demographic structure of the POP study cohort, which reduces the potential for noise and bias.

The evidence supporting these associations could be strengthened further by additional studies. In particular, the BiB study did not have a 36wkGA sample. A key feature of the POP study was the availability of a blood sample near term. This timing of blood sampling was purposeful as we had previously suggested that testing in late pregnancy is likely to yield the strongest prediction of complications at term[12]. Hence, it would be interesting to know whether the stronger associations near term are also observed in other cohorts. The current study also indicated that the combination of EFW<20th and metabolite ratio >80th might be optimal for screening near term. The combination of both tests positive had a positive LR of 11, >50% sensitivity and a 5% false positive rate. Similarly, when both tests were normal, the negative LR was 0.07 and the NPV was 99.7%. Hence, two thirds of women were identified as being either at very high risk or very low risk of FGR. However, these thresholds were not pre-specified and further validation would be informative. Importantly, validation studies will not need to perform serial ultrasound to define FGR using fetal growth velocity, as the ratio was strongly associated with FGR when the outcome was simply defined on the basis of birth weight <3rd percentile (**Fig. 2c, Supplementary Table 9)**. Finally, the study design was focused on identifying novel predictors of term FGR, and the rationale is described above. Although we used preterm blood samples in both the initial phase of biomarker discovery and external validation, metabolites were selected on the basis of their association with term FGR. The metabolite ratio at 28wkGA was only weakly associated with the risk of preterm delivery of an FGR infant in the POP study (**Extended Data Fig. 7**). This observation supports the view that that the

pathophysiology of preterm and term FGR are dissimilar[26]. Further studies could use the same methods employed in the present analysis to identify novel predictors for preterm FGR.

One limitation of the present study is that levels of metabolites were assessed using relative concentrations rather than absolute units. Further studies could also address quantification of the absolute concentration of metabolites and further analysis of inter- and intra-assay variability. However, it is also likely that, even if the absolute concentrations had been known, we would have expressed metabolites in terms of MoMs, as the use of relative concentrations in pregnancy is already commonplace in other screening contexts. For example, in Down syndrome screening maternal levels of proteins and hormones are expressed as MoMs adjusted for gestational age and maternal characteristics[27]. Development of local reference ranges has been shown to increase the predictive value of the tests[28]. The use of MoMs is thought to remove site to site variation in levels (in essence, batch effects), hence the MoM more closely reflects the biological variation rather than technical variation in the processing of samples or the analysis platform employed. Similarly, in the present study, analysis of metabolites by study and sample specific MoMs may have contributed to external validation in the BiB cohort, overcoming differences in population, sample type, gestational age and sample processing.

Given the associations, we believe that the metabolite ratio, when combined with ultrasonic assessment of fetal weight, has potential as a screening test. Next steps are development of quantitative assays, validation of diagnostic effectiveness in further populations, and further assessment of how the metabolites can be used in combination with other biomarkers, such as sFLT1 and PlGF. We found (**Fig. 2c)** that the metabolite ratio was much more strongly associated with FGR in the absence of preeclampsia than the sFLT1:PlGF ratio and the converse was also observed. Consequently, when we repeated the analysis and employed a definition of FGR which included both maternal preeclampsia and neonatal morbidity[7] both ratios were strongly predictive

when combined with ultrasound (LR+ 15 to 20) (**Supplementary Table 10**). Considering the complementary nature of the two ratios, combining proteins and metabolites into a single ratio may be considered but is beyond the scope of the present study.

Confirmation of the generalizable diagnostic effectiveness of the metabolite ratio would provide a rationale for trials of screening and intervention. Screening using the metabolite ratio could take different forms. It could be used to select women for ultrasound or, alternatively, both scan and the metabolite ratio could be assessed at the same visit. Moreover, the metabolite ratio worked similarly irrespective of maternal BMI, hence it could be used to identify women who require further assessment if a false negative ultrasound is more likely (e.g. in super obese women). Finally, the ratio may be especially useful for screening low risk women. Generally, when diagnostic testing is applied to high risk women, the negative predictive value of the test is prioritised as the major concern is falsely reassuring women who actually have a problem. However, when screening low risk women, the major concern is false positives, as these can lead to medicalisation of healthy women and iatrogenic harm. Hence, the potential clinical utility of the metabolite ratio is underlined by the fact that the combination of ultrasound and the ratio yielded high positive LRs when both tests were positive and very low negative LRs when both tests were negative.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at [link to the online version of the paper].

Acknowledgements

The work was supported by the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre (Women's Health theme), the Medical Research Council (United Kingdom; 1100221 to G.C.S.S. and D.S.C.-J., MR/N024397/1 to D.A.L.), the Wellcome Trust (WT101597MA), National Institutes of Health (R01 DK10324), the European Research Council (669545), and the NIHR Biomedical Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol (Reproductive and Perinatal Health theme), which funds N.M.'s PhD studentship. N.G., N.M. and D.A.L. work in a unit that receives support from the MRC (MC_UU_00011/6) and University of Bristol. The funders did not have any role in the design, analysis or preparation of the manuscript for publication. We are grateful to the participants in the POP and Born in Bradford studies and staff who recruited and assessed these participants. We would like to thank L. Bibby, S. Ranawaka, K. Holmes, J. Gill and R. Millar for technical assistance.

Author contributions

G.C.S.S. had the original idea. G.C.S.S., D.S.C.-J. and D.A.L. designed the experiments. U.S. and G.C.S.S. conceived the analysis. N.G., N.M. and U.S. conducted the analysis. E.C., F.G. and D.S.C.-J. conducted the laboratory work. U.S. and G.C.S.S. drafted the initial version of the MS. All authors have seen and approved the final version of the MS.

Competing interests

Direct: Cambridge Enterprise (UK) have filed a patent relating to the associations described in this paper with U.S., D.S.C.-J. and G.C.S.S. as the named inventors.

Indirect: G.C.S.S. reports research support in kind from GE and Roche, and financial support of research from GSK and Sera Prognostics. G.C.S.S. has been paid to attend advisory boards by GSK and Roche. G.C.S.S. has acted as a paid consultant to GSK and is a member of a Data Safety and Monitoring Committee for a GSK vaccine trial. D.A.L. has received support in kind from Roche Diagnostics and Medtronic Ltd.

Additional information

Supplementary information is available for this paper at [link to Supplementary Information].

Correspondence and request for materials should be addressed to U.S. or G.C.S.S.

References

1. Gardosi, J., Madurasinghe, V., Williams, M., Malik, A. & Francis, A. Maternal and fetal risk factors for stillbirth: population based study. *BMJ* **346**, f108 (2013).
2. Chauhan, S.P*., et al.* Neonatal Morbidity of Small- and Large-for-Gestational-Age Neonates Born at Term in Uncomplicated Pregnancies. *Obstet. Gynecol* **130**, 511-519 (2017).
3. Moraitis, A.A., Wood, A.M., Fleming, M. & Smith, G.C.S. Birth Weight Percentile and the Risk of Term Perinatal Death. *Obstetrics and Gynecology* **124**, 274-283 (2014).
4. MacKay, D.F., Smith, G.C., Dobbie, R. & Pell, J.P. Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren. *PLoS. Med* **7**, e1000289 (2010).
5. Kallen, B., Finnstrom, O., Nygren, K.G. & Otterblad Olausson, P. Association between preterm birth and intrauterine growth retardation and child asthma. *Eur Respir J* **41**, 671-676 (2013).
6. Barker, D.J. Adult consequences of fetal growth restriction. *Clin. Obstet. Gynecol* **49**, 270-283 (2006).
7. Gaccioli F, S.U., Cook E, Hund M, Charnock-Jones DS, Smith GCS. Screening for fetal growth restriction using ultrasound and the sFLT1/PlGF ratio in nulliparous women: a prospective cohort study. *Lancet Child Adolesc Health* **2**, 569-581 (2018).
8. Reddy, U.M*., et al.* Prepregnancy risk factors for antepartum stillbirth in the United States. *Obstet. Gynecol* **116**, 1119-1126 (2010).
9. Monier, I*., et al.* Poor effectiveness of antenatal detection of fetal growth restriction and consequences for obstetric management and neonatal outcomes: a French national study. *BJOG* **122**, 518-527 (2015).
10. Bricker, L., Medley, N. & Pratt, J.J. Routine ultrasound in late pregnancy (after 24 weeks' gestation). *Cochrane. Database. Syst. Rev* **6**, CD001451 (2015).

11. RCOG. Guideline No. 31: The investigation and management of the small for gestational age fetus. 1-34 (London, UK, 2013).

12. Smith, G.C. Researching new methods of screening for adverse pregnancy outcome: lessons from pre-eclampsia. *PLoS Med* **9**, e1001274 (2012).

13. Spong, C.Y. Defining "term" pregnancy: recommendations from the Defining "Term" Pregnancy Workgroup. *JAMA* **309**, 2445-2446 (2013).

14. MacDonald, T.M., McCarthy, E.A. & Walker, S.P. Shining light in dark corners: diagnosis and management of late-onset fetal growth restriction. *Aust N Z J Obstet Gynaecol* **55**, 3-10 (2015).

15. Sovio, U., White, I.R., Dacey, A., Pasupathy, D. & GCS, S. Screening for fetal growth restriction with universal third trimester ultrasonography in nulliparous women in the Pregnancy Outcome Prediction (POP) study: a prospective cohort study. *Lancet* **386**, 2089-2097 (2015).

16. Gong, S*., et al.* Placental polyamine metabolism differs by fetal sex, fetal growth restriction, and preeclampsia. *JCI Insight* **3**(2018).

17. Sharp, S.J., Poulaliou, M., Thompson, S.G., White, I.R. & Wood, A.M. A review of published analyses of case-cohort studies and recommendations for future reporting. *PLoS One* **9**, e101176 (2014).

18. Zeisler, H*., et al.* Predictive Value of the sFlt-1:PlGF Ratio in Women with Suspected Preeclampsia. *N. Engl. J. Med* **374**, 13-22 (2016).

19. Delaney, S.S., Coley, R.Y. & Brown, Z. 1,5-Anhydroglucitol: a new predictor of neonatal birth weight in diabetic pregnancies. *Eur J Obstet Gynecol Reprod Biol* **189**, 55-58 (2015).

20. Eriksson, H., Gustafsson, J.A. & Sjovall, J. Excretion of steroid hormones in adults. C19 and C21 steroids in faeces from pregnant women. *Eur J Biochem* **12**, 520-526 (1970).

21. Tanaka, T., Suguro, N. & Kubodera, A. A simple radioimmunoassay for estriol 3-sulfate in pregnancy plasma without deconjugation. *Steroids* **46**, 649-657 (1985).

22. Kunz, J. & Keller, P.J. Ultrasound and biochemical findings in intrauterine growth retardation. *J Perinat Med* **4**, 85-94 (1976).

23. Raeside, J.I. A Brief Account of the Discovery of the Fetal/Placental Unit for Estrogen Production in Equine and Human Pregnancies: Relation to Human Medicine. *Yale J Biol Med* **90**, 449-461 (2017).

24. Brien, M., Berthiaume, L., Rudkowska, I., Julien, P. & Bilodeau, J.F. Placental dimethyl acetal fatty acid derivatives are elevated in preeclampsia. *Placenta* **51**, 82-88 (2017).

25. Brosens, I., Pijnenborg, R., Vercruysse, L. & Romero, R. The "Great Obstetrical Syndromes" are associated with disorders of deep placentation. *Am J Obstet Gynecol* **204**, 193-201 (2011).

26. Mifsud, W. & Sebire, N.J. Placental pathology in early-onset and late-onset fetal growth restriction. *Fetal Diagn. Ther* **36**, 117-128 (2014).

27. Neveux, L.M., Palomaki, G.E., Larrivee, D.A., Knight, G.J. & Haddow, J.E. Refinements in managing maternal weight adjustment for interpreting prenatal screening results. *Prenat. Diagn* **16**, 1115-1119 (1996).

28. Sorensen, S., Momsen, G., Sundberg, K., Friis-Hansen, L. & Jorgensen, F.S. First-trimester risk calculation for trisomy 13, 18, and 21: comparison of the screening efficiency between 2 locally developed programs and commercial software. *Clin Chem* **57**, 1023-1031 (2011).

Figure legends

Fig. 1. **Levels of predictive metabolites at four gestational time points.** (**a**-**d**) Mean (95% CI) relative concentrations of the four selected metabolites in maternal serum across gestation in cases of fetal growth restriction (FGR) born at term or non-cases born at term (control) in the Pregnancy Outcome Prediction (POP) study: 1-(1-enyl-stearoyl)-2-oleoyl-GPC (**a**), 1,5-anhydroglucitol (**b**)**,** 5α-androstan-3α,17α-diol disulfate (**c)**, and N1,N12-diacetylspermine (**d**). The numbers of control/case samples were 278/171 at 12 wkGA, 284/171 at 20 wkGA, 283/169 at 28 wkGA, and 275/162 at 36 wkGA. Metabolites were quantified using area-under-the-curve of primary MS ions and expressed as the multiple of the median value for all batches processed on a given day (see Methods). Term FGR was defined as delivery at ≥37 weeks of gestational age with customized birthweight <3rd percentile, or <10th percentile with abdominal circumference growth velocity in the lowest decile (see Methods). The *P* values for the interaction between wkGA and FGR from the mixed effects regression models are listed in **Supplementary Table 2**. The *P* values for the effect of advancing gestational age between 12 wkGA and 28 wkGA were <0.0001 for all four metabolites.

Fig. 2. Receiver operating characteristic (ROC) curve analyses for the prediction of fetal growth restriction (FGR). (**a**) The metabolite ratio (solid line) and the sFLT1:PlGF ratio (broken line) at 36 weeks of gestational age (wkGA) comparing term FGR (n=162) and controls (n=275) (area under the ROC curve [AUC] [95% CI] = 0.78 [0.73 to 0.82] and 0.64 [0.58 to 0.69], respectively, DeLong test *P*=0.0001 [two-sided] for the AUC comparison). The diagonal line represents the AUC of 0.5 (= no discrimination). (**b**) The metabolite ratio in fasting maternal plasma at 24-28 wkGA from the Born in Bradford (BiB) study samples 1 and 2 (BiB 1 and BIB 2, respectively) in all subsequent FGR cases (n=20 and 41 in BIB 1 and BIB 2, respectively) and controls (n=950 and 1513 in BIB 1 and BIB2, respectively) (BiB 1 AUC = 0.68 [95% CI: 0.55 to 0.81], *P*=0.0029 [one-sided]; BiB 2 AUC = 0.62 [0.54 to 0.71], *P*=0.0018 [one-sided]; BiB 1 & 2 AUC = 0.64 [0.57 to 0.71], *P*<0.0001 [one-sided]). Also

shown is the metabolite ratio in non-fasting maternal serum at 20, 28 and 36 wkGA from the POP study in all subsequent FGR ≥28 wkGA cases (n=141, 136 and 117, respectively) and controls (n=295, 294 and 281, respectively) (20 wkGA AUC = 0.64 [95% CI: 0.58 to 0.69], *P*<0.0001; 28 wkGA AUC = 0.72 [95% CI: 0.67 to 0.77], *P*<0.0001; 36wk GA AUC = 0.80 [95% CI: 0.75 to 0.85], *P*<0.0001). (**c**,**d**) The metabolite ratio and the sFLT1:PlGF ratio at 36 wkGA in relation to term FGR by phenotype (**c**), and the metabolite ratio at 36 wkGA in relation to term FGR by maternal or fetal characteristics (**d**). In **c** and **d**, the total number of FGR cases was 162 and the total number of controls was 275. In **c**, BW<3rd percentile n=110; BW 3rd to <10th percentile + ACGVD1 n=52; preeclampsia n=14; no preeclampsia n=148, and the dotted line represents the AUC of 0.5 (= no discrimination). In **d**, weight categories were based on body mass index cut-offs of 25 and 30 kg/m^2. Underweight women (<18.5 kg/m^2, n=9) were included in the normal weight group. The analysis of estimated fetal weight (EFW) included 160 cases and 273 controls due to missing values. In **c** and **d**, the vertical dashed lines represent AUC comparing all cases and controls. In **a, c** and **d**, term FGR was defined as delivery at ≥37 weeks of gestational age with customized birthweight (BW) <3rd percentile, or <10th percentile with abdominal circumference growth velocity in the lowest decile (ACGVD1, see Methods). In **b**, FGR was defined as subsequent delivery with birthweight <3rd centile corrected only for GA and fetal sex (see Methods and Supplementary Information). The metabolite ratio was calculated using the multiple of the median values, as the ratio of the product of 1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1) and 1,5-anhydroglucitol (1,5-AG) divided by the product of 5alpha-androstan-3alpha,17alpha-diol disulfate and N1,N12-diacetylspermine.

Table 1. Metabolite measurements and their products and ratios at 36 wkGA in relation to FGR[a] at term.

Metabolite	AUC[b] (95% CI)	*P*[c]	Odds ratio[d] (95% CI)		
			Unadjusted	Adjusted[e]	Fully adjusted[f]
(A) 1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1)	0.64 (0.58 to 0.69)	$2.3x10^{-7}$	1.76 (1.41 to 2.21)	1.68 (1.34 to 2.12)	1.76 (1.38 to 2.23)
(B) 1,5-anhydroglucitol (1,5-AG)	0.65 (0.60 to 0.71)	$8.4x10^{-7}$	1.79 (1.40 to 2.27)	1.65 (1.29 to 2.11)	1.62 (1.26 to 2.07)
(C) 5α-androstan-3α,17α-diol disulfate	0.69 (0.64 to 0.74)	$3.2x10^{-11}$	0.51 (0.41 to 0.64)	0.52 (0.42 to 0.65)	0.51 (0.41 to 0.64)
(D) N1,N12-diacetylspermine	0.66 (0.61 to 0.71)	$1.1x10^{-5}$	0.63 (0.51 to 0.79)	0.58 (0.45 to 0.74)	0.58 (0.45 to 0.74)
Numerator of the metabolite ratio (A x B)	0.70 (0.64 to 0.75)	$7.9x10^{-11}$	2.18 (1.70 to 2.81)	2.02 (1.56 to 2.61)	2.01 (1.55 to 2.61)
Denominator of the metabolite ratio (C x D)	0.71 (0.66 to 0.76)	$7.6x10^{-12}$	0.49 (0.39 to 0.62)	0.48 (0.38 to 0.60)	0.47 (0.37 to 0.60)
Metabolite ratio (A x B) / (C x D)	0.78 (0.73 to 0.82)	$1.1x10^{-21}$	2.93 (2.25 to 3.80)	2.86 (2.20 to 3.73)	2.82 (2.17 to 3.68)

The total number of women who had metabolite measurements at 36 wkGA was 437, including 162 cases of FGR and 275 controls born at term. [a]FGR at term was defined as delivery at ≥37wkGA with customized birth weight <3rd percentile, or customized birth weight <10th percentile with abdominal circumference growth velocity in the lowest decile (see Methods). [b]AUC was based on the metabolites alone. [c]Calculated from linear regression using the Wald test, with the null hypothesis that the coefficient = 0. [d]Odds ratios were given for one standard deviation higher value of the log-transformed metabolite, product or ratio. [e]Adjusted for the log-transformed sFlt-1:PlGF ratio at 36wkGA. [f]Additionally adjusted for maternal age (linear and quadratic term) and body mass index at 12wkGA. wkGA, weeks of gestational age; FGR, fetal growth restriction; CI, confidence interval; AUC, area under the ROC curve.

Table 2. Diagnostic effectiveness of ultrasonic and biochemical screening at 36 wkGA for delivery of an infant with FGR[a] at term.

Screening test	TP/FP	TN/FN	Screen^{+} Comp	Positive LR (95% CI)	Negative LR (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV[c] (95% CI)	NPV[c] (95% CI)	DOR (95% CI)
Ultrasonic EFW <10th	110/39	234/50	17.1	4.8 (3.5-6.6)	0.36 (0.29-0.46)	68.8 (61.1-75.5)	85.7 (81.0-89.4)	18.0 (13.2-24.1)	98.4 (97.8-98.8)	13.2 (8.2-21.0)
sFLT1:PlGF ratio >38	49/33	240/111	12.5	2.5 (1.7-3.8)	0.79 (0.71-0.88)	30.6 (23.9-38.3)	87.9 (83.5-91.3)	10.4 (6.9-15.4)	96.5 (95.7-97.2)	3.2 (1.9-5.3)
[b]Metabolite ratio >85th	86/37	236/74	15.0	4.0 (2.8-5.5)	0.54 (0.45-0.64)	53.8 (45.9-61.4)	86.4 (81.8-90.0)	15.3 (10.9-21.1)	97.6 (96.9-98.2)	7.4 (4.7-11.9)
Ultrasonic EFW <10th and sFLT1:PlGF ratio >38	34/6	267/126	3.1	9.7 (4.2-22.5)	0.81 (0.74-0.87)	21.3 (15.5-28.4)	97.8 (95.2-99.0)	30.6 (15.1-52.2)	96.5 (95.7-97.1)	12.0 (4.9-27.4)
Ultrasonic EFW <10th and metabolite ratio >85th	56/4	269/104	3.1	23.9 (8.8-64.6)	0.66 (0.59-0.74)	35.0 (27.9-42.8)	98.5 (96.1-99.5)	52.1 (27.7-75.6)	97.1 (96.4-97.7)	36.2 (13.7-94.9)

The total number of women in this analysis was 433, including 160 cases of FGR and 273 controls, due to missing values in EFW for two cases and two controls. [a]FGR at term was defined as delivery at ≥37wkGA with customised birth weight <3rd percentile, or customised birth weight <10th percentile with abdominal circumference growth velocity in the lowest decile (see Methods). [b]Metabolite ratio is the ratio of two products of metabolites (see Methods). As the sFLT1:PlGF ratio >38 approximates to the 85th percentile in the whole POP study cohort, we selected the same threshold in this analysis. [c]Due to the case-cohort design, the proportion of screen positives was calculated in the random subcohort, i.e. comparator group, in women who had all three measurements (EFW, sFLT1:PlGF, metabolite ratio) available (n=287 including 14 cases of FGR and 273 non-cases), and PPV and NPV were weighted by the inverse of the random subcohort sampling fraction. The proportion of screen positives, sensitivity, specificity, PPV and NPV are given in percentages (%). wkGA, weeks of gestational age; FGR, fetal growth restriction; TP, true positive; FP, false positive; TN, true negative, FN, false negative; Screen^{+}, screen positive; Comp, comparator group; LR, likelihood ratio; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; DOR, diagnostic odds ratio; EFW, estimated fetal weight; sFLT1, soluble fms-like tyrosine kinase 1; PlGF, placenta growth factor.

Methods

The approach of the study was as follows: (i) identify candidate metabolite predictors using the 12, 20 and 28wkGA POP study samples, (ii) validate predictors using the 36wkGA POP study sample and identify those which were predictive of term FGR independently of maternal characteristics and the sFLT-1:PlGF ratio, (iii) validate the predictors externally using the Born in Bradford (BiB) cohort.

Study design

The POP study has been described previously in detail[29,30]. It was a prospective cohort study of unselected nulliparous women with a singleton pregnancy attending the Rosie Hospital, Cambridge, UK, between Jan 2008 and Jul 2012. Participants had repeated blood sampling and fetal biometry at 12, 20, 28 and 36wkGA. Outcome data were obtained by linkage to the hospital's electronic databases and individual review of paper case records. Ethical approval was obtained from the Cambridgeshire 2 Research Ethics Committee (reference number 07/H0308/163). All study participants gave written informed consent. This study is reported according to the STARD 2015 guidelines for reporting diagnostic accuracy studies (http://www.stard-statement.org/). A case-cohort design within the POP study was used for the metabolomics analysis[31]. FGR at term was defined as delivery at ≥37wkGA and (a) customized birth weight (BW) centile <3rd or (b) customized[32] birth weight (BW) centile <10th combined with abdominal circumference growth velocity (ACGV) in the lowest decile between 20 and 36 wkGA. A random sample of the cohort was selected as a comparison group

External validation: the BiB Cohort

The BiB cohort has been described in detail elsewhere[33,34] and was conducted between 2007 and 2011. The cohort members, excluding those with pre-existing diabetes, were invited for a glucose tolerance test at 24-28wkGA, and 85% of the women had valid test data. 1000 women from the BiB cohort were selected randomly from this group including women who had stored fasting plasma. BiB

also had a second sample of 2000 women who underwent metabolic profiling at Metabolon. Whilst the first BiB sample (n=1000) was random, the second sample selection was designed similarly to POPs in a case-cohort design, sampled for cases of gestational diabetes, gestational hypertension, pre-eclampsia, preterm birth and still birth, but not FGR. Therefore, we used the FGR cases and non-cases available in the random sub-cohort (n=1554) from the second BiB sample in the analysis. This was an independent sample to the first BiB sample and we did not pool the metabolite MoMs from the two samples due to different normalisation. Both BiB samples are half White British and half Pakistani origin because these are the main homogeneous ethnic groups in the area.

Ultrasonic measurements of abdominal circumference and fully customized BW centiles were not available for all women in the BiB cohort. Therefore, a modified definition of FGR was employed in the BiB cohort: birth weight <3rd centile corrected for GA and fetal sex using the Hadlock 1991[35] reference range (see Analysis plan below). When we compared data from the POP study and BiB, the same modified definition of FGR was employed in the POP study to allow a direct comparison between the two cohorts. *P* values for external validation were one-sided as validation was directional, i.e. we would not regard an association as validated if the *P* value was below the given threshold but the association was in the opposite direction to that predicted. The analysis plan pre-specific logistic regression as the analytic method to generate the *P* value. We also performed a post hoc analysis of the BiB cohort using ROC curve analysis to allow direct comparison of associations with the POP study. Again, a one-sided *P* value was reported.

Biochemical analyses

Measurement of sFLT1 and PlGF protein levels (undertaken only in the POP study) was performed on maternal serum using the Roche Cobas e411 immunoassay platform, as previously described[36]. Metabolomic analysis was performed by Metabolon (Research Triangle Park, NC, USA), blinded to the patients' clinical information and pregnancy outcome, as previously described[31]. In both studies, Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS) was

used as the analysis platform[37]. Metabolite concentrations were quantified using area-under-the-curve of primary MS ions and were expressed as the multiple of the median (MoM) value for all batches processed on the given day. In the POP study, analysis batches contained 36 maternal serum samples each and all samples from the same woman were included in the same batch (hence including the full range of gestational ages). Moreover, batches were designed so that the proportion of samples from cases and controls was the same in all batches. The calculated metabolite products and ratios were derived from multiplication and division of the MoM values. 1193 untargeted metabolites were measured from each sample, 837 of known structural identity. Eight xenobiotic metabolites were not analysed as they demonstrated minimal variation. In the POP study, metabolomics was performed on serial non-fasting serum obtained at around 12, 20, 28 and 36wkGA, whereas BiB samples were plasma samples (ethylenediamine tetraacetic acid tubes) obtained once from each woman at 24-28wkGA after an overnight fast. The relative standard deviations (%RSD) for the four metabolites used in the metabolite ratio were: 19.5% for 1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1), 10.3% for 1,5-anhydroglucitol (1,5-AG), 5.8% for 5alpha-androstan-3alpha,17alpha-diol disulfate and 10.8% for N1,N12-diacetylspermine. These values were derived from the QC matrix of pooled EDTA plasma or serum.

Statistical analysis

The calculation of metabolite products and ratios was performed using the metabolite MoMs. Metabolite MoMs, products and ratios were log-transformed prior to linear regression. Additionally, in logistic regression, the log transformed values were converted to z scores to allow direct comparison of the estimated effect sizes. Initial selection of predictive metabolites involved fitting longitudinal linear mixed models for each metabolite using measurements from 12wkGA, 20wkGA, and 28wkGA, to generate a difference in the metabolite means and associated *P* value in the maternal serum at 20wkGA and/or 28wkGA (composite Chi-squared test) comparing term FGR cases and controls. We included interaction terms between term FGR and gestational age to identify

differences and the metabolites were then ranked by the composite *P* value at 20/28wkGA. Excess of low *P* values was tested using a one-sample Kolmogorov-Smirnov test against the theoretical uniform distribution of *P* values between 0 and 1. The 100 metabolites with the lowest *P* values were selected for further study. Internal validation used the 36wkGA sample in the same women and cases and controls were compared using linear regression. Internal validation was accepted if the *P* value at 36wkGA was below the Bonferroni-corrected threshold $P<5\times10^{-4}$. Forward-stepwise logistic regression ($P<0.05$ for entry and $P<0.1$ for removal) was used to select independent predictors of term FGR. In addition to the metabolites internally validated at 36wkGA, the forward-stepwise logistic regression included the sFlt-1:PlGF ratio at 36wkGA, maternal age (linear and quadratic terms) and maternal BMI at 12wkGA. The metabolites selected based on the forward-stepwise logistic regression were further assessed on whether they improved the area under the receiver operating characteristic (ROC) curve (AUC) in the prediction of term FGR over the sFLT1:PlGF ratio. In this step, the AUC was estimated using 1000-fold bootstrapping to avoid optimism through overfitting. The metabolites were added into a logistic regression model for the sFLT1:PlGF ratio, starting from the most informative metabolite, until the increase in the corrected AUC on adding an additional metabolite was <0.01. The metabolites from this step were then used to calculate products and ratios of the unprocessed MoMs generated by Metabolon and AUCs (95%CI) of the products and ratios were reported. These were not corrected for optimism as they were treated as single predictors in the analyses and this did not involve fitting coefficients using a multivariable model. Additionally, unadjusted and adjusted odds ratios (95% CI) were reported for a one standard deviation higher value of the log-transformed metabolite, product or ratio. Standard screening test statistics (sensitivity, specificity, positive and negative likelihood ratio, positive and negative predictive value and diagnostic odds ratio) were calculated from 2x2 tables in the POP study cohort, weighting the comparison group by the inverse of the sampling fraction where appropriate. A power calculation for validation of the metabolite ratio was performed using the effect size obtained from the POP study for the 28wkGA sample and the same FGR definition that was employed in the BiB

study. External validation in the BiB study was pre-specified in an analysis plan (see below) which was informed by the power calculation. To account for differences in the two samples from the BiB study, the pooled statistics (odds ratio and AUC and their 95%CI) were obtained by first taking a z score of the log-transformed ratio separately in both samples and by calculating the statistics from the pooled data of sample-specific z scores. Statistical analysis was performed using Stata version 15.1 and R version 3.4.4.

Analysis plan for external validation of associations between metabolites and FGR in the Born in Bradford study

Outcome

FGR, defined as birth weight percentile <3rd, applying the 1991 Hadlock formula[35] to sex-adjusted weights (see Methods section below). Births at any gestational age subsequent to the measurement of the metabolites are included.

Exposures

Scaled imputed metabolite values (multiples of the median) from maternal serum or plasma are used to calculate the following from the measurements taken at 24-28 weeks of gestation:

The main exposures are:

1. the ratio of two products of metabolites: **(1. x 2.) / (3. x 4.)**, where

1. 1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1)

2. 1,5-anhydroglucitol (1,5-AG)

3. 5alpha-androstan-3alpha,17alpha-diol disulfate

4. N1,N12-diacetylspermine

2. 5alpha-androstan-3alpha,17alpha-diol disulfate (as sole predictor)

Note: we will accept validation as a Bonferroni corrected threshold for alpha (<0.025), one-sided (given known directionality of association being tested).

Secondary exposures are

(i) The product of metabolites **3.** and **4.** above, i.e. the denominator of main exposure 1.

(ii) The product of metabolites **1.** and **2.** above, i.e. the numerator of main exposure 1.

(iii) Steroid ratio / Polyamine ratio, where

Steroid ratio = 4-androsten-3beta,17beta-diol monosulfate (2) / 5alpha-androstan-3alpha,17alpha-diol disulfate.

Polyamine ratio = N1,N12-diacetylspermine / Acisoga.

(iv) All of the other individual metabolites listed above (metabolites **1.**, **2.**, **4.**, 4-androsten-3beta,17beta-diol monosulfate (2) and Acisoga) as sole predictors.

Given the number of hypotheses, these will be treated as "hypothesis generating" and accepting an uncorrected $P<0.05$ as the metabolite being potentially associated with FGR but requiring further validation. However, if any of the secondary exposures are $P<0.00625$ (one-sided, Bonferroni corrected for 8 comparisons – two products of metabolites, one ratio of ratios and five individual measures), we would accept this as validation of the POP study results.

Any associations observed in the opposite direction from the POP study will be disregarded, given the use of one-sided tests.

Transformation of exposures

Log-transform all exposures, e.g. if the main exposure variable 1 is named mainratio, log10mainratio = log10(mainratio).

Turn the log-transformed ratios into z scores. In the POPs, the mean and SD of log-transformed ratios for calculating z scores were obtained from the comparator group which is representative of the whole POPs cohort. In the BiB study, you can use the population mean and SD, log10mainratioZ = [log10mainratio - mean(log10mainratio)] / SD(log10mainratio).

Statistical analysis

Fit a logistic regression model separately for each exposure and FGR (outcome). Report the odds ratio, 95%CI and *P* value (one-sided) from each analysis. Perform a ROC curve analysis and calculate AUC (95% CI) for each exposure.

Methods for calculating gestational age and fetal sex adjusted birth weight percentiles

There were 13524 participants in the BiB dataset with information on birth weight, fetal sex and gestation length (both in weeks and days and in completed weeks). We adjusted each of these birth weights for fetal sex, applied the Hadlock 1991 formulas to these sex-adjusted weights and defined FGR as follows:

1. Participants were grouped by gestation length (in completed weeks). To get adequate numbers of participants in each group (>50), we combined the weeks 24-28, 29-31, 32-33 and 42-44. All other weeks (34-41) were analysed independently, so that there were 12 groups altogether.
2. Within each group *i* we calculated the mean birth weights for both males ($m[i]$) and females ($f[i]$), and the difference ($d[i]$) in the means for each group ($d[i]= m[i]- f[i]$).
3. We adjusted the birth weights within each group as follows:

$$m^*[i] = m[i] - \frac{1}{2}d[i],$$

$$f^*[i] = f[i] + \frac{1}{2}d[i],$$

where $m^*[i]$ and $f^*[i]$ are the sex-adjusted birth weights.

4. We applied the 1991 Hadlock formulas to each of the 13524 participants (using gestation length as weeks and days [in decimal form]). We then calculated z-scores for each participant using the sex-adjusted birth weights defined above.
5. We defined FGR in the BiB dataset as a z-score (defined above) <3rd percentile of all participants.

A similar method in the POP study cohort (n=4212) was applied to obtain gestational age and fetal sex adjusted birth weight percentiles and to define FGR. This definition was used when the results from the BiB and POP study cohorts were presented together.

Data availability

Source data for Figs. and Extended Data Figs. are available online. Since the individual patient data contain confidential information, it can be supplied only in an anonymised format to suitably qualified researchers who can make appropriate institutional commitments relating to data security and confidentiality. Data requests should be addressed to U.S. or G.C.S.S.

References

29. Pasupathy, D.*, et al.* Study protocol. A prospective cohort study of unselected primiparous women: the pregnancy outcome prediction study. *BMC. Pregnancy. Childbirth* **8**, 51 (2008).

30. Gaccioli, F., Lager, S., Sovio, U., Charnock-Jones, D.S. & Smith, G.C.S. The pregnancy outcome prediction (POP) study: Investigating the relationship between serial prenatal ultrasonography, biomarkers, placental phenotype and adverse pregnancy outcomes. *Placenta* **59**, S17-S25 (2017).

31. Sovio, U.*, et al.* 4-Hydroxyglutamate is a novel predictor of pre-eclampsia. *Int J Epidemiol* (2019).

32. Gardosi, J., Mongelli, M., Wilcox, M. & Chang, A. An adjustable fetal weight standard. *Ultrasound Obstet. Gynecol* **6**, 168-174 (1995).

33. Raynor, P. & Born in Bradford Collaborative, G. Born in Bradford, a cohort study of babies born in Bradford, and their parents: protocol for the recruitment phase. *BMC Public Health* **8**, 327 (2008).

34. Wright, J.*, et al.* Cohort Profile: the Born in Bradford multi-ethnic family cohort study. *Int J Epidemiol* **42**, 978-991 (2013).

35. Hadlock, F.P., Harrist, R.B. & Martinez-Poyer, J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology* **181**, 129-133 (1991).

36. Sovio, U.*, et al.* Prediction of Preeclampsia Using the Soluble fms-Like Tyrosine Kinase 1 to Placental Growth Factor Ratio: A Prospective Cohort Study of Unselected Nulliparous Women. *Hypertension* **69**, 731-738 (2017).

37. Evans, A.M., DeHaven, C.D., Barrett, T., Mitchell, M. & Milgram, E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem* **81**, 6656-6667 (2009).