

**Automated assessment of the first trimester placental volume and uterine
vasculature using three-dimensional power Doppler ultrasound to predict
adverse perinatal outcomes**

Sam Mathewlynn

BSc (Hons), MBChB, MRCOG

St. Anne's College



A thesis submitted for the degree of Doctor of Philosophy
to the Faculty of Medical Sciences, University of Oxford.

Hilary Term

January 2026

TABLE OF CONTENTS

Abstract.....	7
Acknowledgements.....	8
List of figures	9
List of equations.....	11
List of tables	12
List of abbreviations	14
[1] Introduction	18
1.1 <i>Clinical motivation</i>	19
1.1.1 Fetal growth restriction	19
1.1.2 Pre-eclampsia	21
1.1.3 Clinical outcomes.....	23
1.2 <i>Pathophysiology.....</i>	24
1.2.1 Uteroplacental haemodynamics and unanswered questions.....	27
1.3 <i>Risk mitigation and the importance of screening</i>	29
1.3.1 Screening approaches	29
1.3.2 Prophylaxis and prevention of harm	30
1.4 <i>First trimester placental biomarkers</i>	31
1.4.1 First trimester placental volume	31
1.4.2 Power Doppler and assessment of placental vascularity	34
1.5 <i>Research aims and hypotheses.....</i>	37
1.6 <i>Publications and presentations arising from this work</i>	38
1.6.1 Publications in peer reviewed journals.....	38
1.6.2 Oral presentations at international conferences.....	39
1.6.3 Poster presentations at international conferences	39
[2] General methods and data sources	40
2.1 <i>Introduction</i>	40
2.1.1 Statement of personal contribution	41

2.2	<i>Methodology for FirstPLUS and OxPLUS</i>	42
2.2.1	Study design	42
2.2.2	Primary and secondary outcomes	43
2.2.3	Ethical approval	43
2.2.4	Eligibility criteria	43
2.2.5	Definitions	44
2.2.6	Pre-eclampsia risk assessment and aspirin use	45
2.2.7	Ultrasound scan protocols	46
2.2.8	Placental scan quality control procedures	49
2.2.9	Statistical methodology	51
2.3	<i>Results</i>	52
2.3.1	FirstPLUS and OxPLUS cohort characteristics	53
2.4	<i>Discussion and conclusion</i>	57
2.4.1	Group comparisons	57
2.4.2	Subgroup comparison	59
2.4.3	Conclusion	61
[3]	Development of centile charts	62
3.1	<i>Introduction</i>	62
3.1.1	Aims	63
3.1.2	Statement of personal contribution	63
3.2	<i>Methodology</i>	64
3.2.1	Study population	64
3.2.2	Adequacy of sample size	64
3.2.3	Exploration of data	65
3.2.4	Models for centile chart construction	65
3.2.5	Evaluation of models	66
3.2.6	Exploration of association with adverse outcomes	67
3.3	<i>Results</i>	68
3.3.1	Cohort characteristics	68
3.3.2	Analysis of distribution	70
3.3.3	Comparison and selection of models	73
3.3.4	Evaluation of centile charts	77
3.3.5	Prediction of adverse outcomes	80
3.4	<i>Discussion and conclusion</i>	82

3.4.1	Comparison with previous work	82
3.4.2	Strengths and limitations	83
3.4.3	Clinical and Research Implications.....	85
3.4.4	Conclusion.....	86
[4]	Development of training and quality control procedures	87
4.1	<i>Introduction</i>	87
4.1.1	Westgard rules	89
4.1.2	Cumulative Sum (CUSUM) methodology.....	90
4.1.3	Summary.....	90
4.1.4	Aims	90
4.1.5	Statement of personal contribution	90
4.2	<i>Methodology</i>	91
4.2.1	Sonographer training protocols for OxPLUS.....	91
4.2.2	Development and delivery of feedback tool.....	92
4.2.3	Delivery of feedback and corrective action.....	95
4.2.4	Multi-rule quality control	96
4.2.5	Cumulative Sum (CUSUM) learning-curve analysis.....	98
4.3	<i>Results</i>	100
4.3.1	Primary training of sonographers	101
4.3.2	Implementation of the OxNNet feedback toolkit.....	102
4.3.3	Multi-rule quality control and CUSUM learning curve analysis.....	105
4.4	<i>Discussion and conclusion</i>	109
4.4.1	Causes of QC failure.....	109
4.4.2	Evaluation of OxNNet feedback toolkit and CUSUM learning curve.....	110
4.4.3	Strengths and limitations	112
4.4.4	Conclusion.....	113
[5]	Prediction of fetal growth restriction and pre-eclampsia	114
5.1	<i>Introduction</i>	114
5.1.1	Aims	115
5.1.2	Statement of personal contribution	115
5.2	<i>Methodology</i>	116
5.2.1	Key datasets and definitions.....	116
5.2.2	Power calculation	116

5.2.3	Exploration of collinearity	116
5.2.4	Exploration of smoothing requirements	116
5.2.5	Selection of variables for final models	117
5.2.6	Validation of models	120
5.2.7	Comparison of models.....	122
5.2.8	Evaluation of model calibration	122
5.3	<i>Results</i>	123
5.3.1	Power calculation	123
5.3.2	Correlation matrix.....	123
5.3.3	Exploration of smoothing requirements	125
5.3.4	Variable selection and model construction	127
5.3.5	Evaluation of model performance.....	129
5.3.6	Evaluation of model calibration	136
5.4	<i>Discussion and conclusion</i>	138
5.4.1	Existing literature	139
5.4.2	Strengths and limitations	142
5.4.3	Conclusion.....	146
[6]	Development of novel markers.....	147
6.1	<i>Introduction</i>	147
6.1.1	Aims	150
6.1.2	Statement of personal contribution	150
6.2	<i>Methodology</i>	151
6.2.1	Sources of data	151
6.2.2	Image processing	151
6.3	<i>Statistical methods</i>	155
6.3.1	Handling of non-normal distributions.....	155
6.3.2	Associations between novel parameters and adverse outcomes.....	155
6.3.3	Subgroup analysis	156
6.4	<i>Results</i>	157
6.4.1	Exploration and transformation of data	158
6.4.2	Associations between novel markers and adverse outcomes.....	163
6.4.3	Subgroup analysis	165
6.5	<i>Discussion and conclusion</i>	166
6.5.1	Findings in the context of current understanding	166

6.5.2	Strengths and limitations	168
6.5.3	Conclusion.....	170
[7]	Reproducibility of OxNNet metrics	171
7.1	<i>Introduction</i>	171
7.1.1	Aims	172
7.1.2	Statement of personal contribution	172
7.2	<i>Methodology</i>	173
7.2.1	Data collection protocol	173
7.2.2	Sample size determination	173
7.2.3	Selection of variables for analysis.....	173
7.2.4	Statistical approach.....	174
7.3	<i>Results</i>	176
7.4	<i>Discussion and conclusion</i>	180
7.4.1	Findings in the context of current understanding	180
7.4.2	Strengths and limitations	181
7.4.3	Conclusion.....	184
[8]	Conclusion.....	185
8.1	<i>Summary of Key Findings</i>	185
8.2	<i>Strengths and Limitations</i>	187
8.3	<i>Areas for Future Development</i>	188
8.4	<i>Conclusion</i>	189
[9]	References.....	190
[10]	Appendices	207

ABSTRACT

Placental insufficiency syndromes, including fetal growth restriction (FGR) and pre-eclampsia (PE), are leading causes of perinatal and maternal morbidity and mortality. Early identification of at-risk pregnancies remains challenging due to limitations in current screening and the complexity of placental biology. This thesis evaluates three-dimensional power Doppler ultrasound, processed using machine learning, for quantifying first trimester placental volume and uterine vasculature with the goal of improving prediction of adverse perinatal outcomes.

Data from the FirstPLUS study were used as a reference population and for model development. The OxPLUS cohort was prospectively established to develop and externally validate novel placental imaging biomarkers and predictive models for placental insufficiency. The OxNNet deep learning framework enabled rapid, reproducible segmentation of the placenta and automated calculation of single vessel fractional moving blood volume (svFMBV).

Prescriptive centile charts for placental volume and svFMBV were constructed, and derived z-scores were significantly associated with FGR and PE. Multivariable models incorporating OxNNet-derived metrics, maternal history, and established biomarkers demonstrated good discrimination for both conditions, which was attenuated but preserved when the models were assessed in the independent OxPLUS cohort. The centile charts themselves have not yet been assessed in an independent population.

This thesis also introduces new quantitative markers of placental and vascular morphology, several of which showed significant associations with adverse outcomes. The feasibility of training sonographers in advanced placental ultrasound was demonstrated, alongside the implementation of structured quality assurance tools, resulting in good reproducibility for most OxNNet-derived metrics.

These findings establish proof of concept for automated placental assessment in early pregnancy and provide a foundation for further research into placental insufficiency. This work advances the quantitative assessment of placental function, but further validation in independent populations, together with assessment of calibration and of clinical impact, would be required before it could inform obstetric practice.

ACKNOWLEDGEMENTS

I am deeply grateful to my supervisors, Prof. Sally Collins and Prof. Antoniya Georgieva. Special thanks to Prof. Collins, whose life's work I have been privileged to build upon. As my main supervisor, mentor, and friend, she has been a constant source of support and inspiration. She created the opportunity for this project and secured the funding that made it possible. I thank the Sir Jules Thorn Trust, whose grant to Prof. Collins enabled OxPLUS and my doctoral work.

To the OxPLUS team, our research midwives and assistants — thank you for helping us reach our ambitious recruitment target. I am indebted to Eleni Fotaki, Ezryah Rzewski, Shanice Akinyombo, and especially Katja Shipp, whose focus and hard work were crucial. I also wish to remember Jess Farmer, missed and held in great fondness. My sincere thanks go to the sonographers at the John Radcliffe Hospital who contributed to the study: Tess Norris, Helen Crichton, Charlotte Drewett, Sian Evans, Shila Sadeghi Kaji, Helen Price, Karolina Edwards, Ali Chevassut, Jayne Lander, Caroline Dillon, Sarah Bendor-Samuel, Roddy Deans, Jane Castle, Joanna Gregg, Ali Fitzpatrick, Beth Leake, Harriet Josephy, Salli Mitchell, Melanie Fenn, Honor Stobart, and Grace Agbo. I am especially grateful to Theo Adu-Bredu, an extraordinary sonographer, prolific researcher, and remarkable colleague. Thanks also to Sam Defreitas, Sandra Helm, Cathrin Shepherd, Donna Morris, and Tayla McMahon, whose contributions kept the study running smoothly.

I would like to acknowledge Charlotte Cose, a gifted poet, for her company, entertainment, Cha-Cha Slide, and endless stickers. My heartfelt thanks also go to Debbie Hedgecote, friend and office companion, and an outstanding study manager whose expertise got the project off the ground and kept things steady through difficult times. And to Mohammadreza Soltaninejad whose software engineering wizardry and machine learning skills made much of this possible. Massive thanks to Prof. Dave Wright and Stephen Gerry for the stimulating statistical discussions, for their advice and for their patience!

Finally, to my wife Riva, and my sons Bodhan and Kashi — thank you for your encouragement, for taking on so much so I could pursue this work, and for believing in me and in us. I know I have been absent during much of this project, and I am profoundly grateful for your patience, love, and unwavering support.

LIST OF FIGURES

Figure 1: The vascular anatomy of the myometrial vascular plexus in the pregnant uterus	26
Figure 2: The position of the five OxNNet-derived 3D-svFMBV locations with respect to the placenta and myometrium	36
Figure 3: FirstPLUS study summary of exclusions and definition of final cohort	52
Figure 4: OxPLUS study summary of exclusions and definition of final cohort	53
Figure 5: Comparison of representation of ethnic categories in the FirstPLUS and OxPLUS cohorts ..	54
Figure 6: Summary of exclusions for centile chart development	68
Figure 7: Paired histograms and quantile-quantile plots for each OxNNet-derived parameter	71
Figure 8: Faceted boxplots for each OxNNet-derived parameter stratified by gestational age in days	72
Figure 9: Final centile charts for placental volume and svFMBV1–5	76
Figure 10: Paired histograms and quantile-quantile plots for z-scores of each OxNNet-derived parameter	79
Figure 11: Forest plots, odds ratios and C-statistics for the prediction of adverse perinatal outcomes in the complete FirstPLUS cohort using z-scores for each OxNNet parameter	81
Figure 12: Example of a diagnostic plot self-assessment tool for nuchal translucency measurements.	88
Figure 13: Summary of sonographers included in training and quality control analyses.....	100
Figure 14: Training events required to achieve competence in A) primary training in first trimester 3D placental ultrasound with power Doppler, and B) further training required in response to the OxNNet feedback toolkit	101
Figure 15: Examples of the self-assessment feedback tool for placental volume (A, B) and svFMBV2 (C, D) showing both positive and negative feedback.....	102
Figure 16: Derivation of OxPLUS subgroup cohorts 1–3	105
Figure 17: Cumulative sum (CUSUM) learning curve analyses for each sonographer performance with respect to subgroup cohort 3 for the entire duration of the OxPLUS study	108
Figure 18: Correlation plot showing associations between candidate variables	124

Figure 19: Plots of placental volume z-score against log relative hazard (with 95% confidence intervals) for each outcome variable, exploring linear modelling and restricted cubic splines with 2 or 3 degrees of freedom.....	126
Figure 20: Receiver operating characteristic (ROC) curves including all models for prediction of A) fetal growth restriction and B) pre-eclampsia	132
Figure 21: Calibration plots for A) the full FGR prediction model (FGR1) and B) the full PE prediction model (PE1)	137
Figure 22: A two-dimensional B-mode ultrasound image of a first trimester pregnancy with three-dimensional OxNNet-derived multiclass segmentation masks overlaid, showing A) the plain two-dimensional ultrasound, B) fetal segmentation, C) placental segmentation, D) amniotic fluid segmentation, and E) complete multiclass segmentation.....	149
Figure 23: Schematic representation of spiral artery jet morphology resulting in A) a low dispersion index and B) a high dispersion index. The solid red line represents the centreline of the vessel jet.	154
Figure 24: Faceted boxplots for novel placental markers plotted against gestational age for A) UPI surface area, B) surface volume ratio, and C) functional spiral artery jet count.....	160
Figure 25: Histograms evaluating the distribution of novel placental markers	161
Figure 26: Histograms demonstrating the results of $\log_{10}$ transformation on distribution of novel markers for A) arcuate artery count, B) radial artery count, and C) FSA jet mean dispersion index..	162
Figure 27: Histograms to evaluate the distribution of z-scores derived from LMS modelling for A) functional spiral artery jet count, B) uteroplacental interface surface area, and C) surface volume ratio	163
Figure 28: Forest plots showing odds ratios for A) fetal growth restriction and B) pre-eclampsia associated with an incremental increase of one for each novel marker	164
Figure 29: Bland-Altman plots for five OxNNet-derived first trimester placental markers demonstrating intra-operator reproducibility (left column) and inter-operator reproducibility (right column).....	179

LIST OF EQUATIONS

Equation 1: Formula to calculate precision in estimation of extreme centiles	64
Equation 2: Bernoulli log-likelihood increment.....	99
Equation 3: Cumulative CUSUM statistic	99
Equation 4: SVR	154
Equation 5: UPI index	154
Equation 6: Dispersion index.....	154

LIST OF TABLES

Table 1: Delphi consensus diagnostic criteria for fetal growth restriction.....	20
Table 2: Comparison of three key diagnostic frameworks for pre-eclampsia.....	22
Table 3: Description of the OxNNet-derived 3D-svFMBV locations	36
Table 4: The major and moderate risk factors for pre-eclampsia as specified in the NICE guideline Hypertension in pregnancy: diagnosis and management (2019)	46
Table 5: Ultrasound machine settings for 3D first trimester placental scan acquisition.....	48
Table 6: Comparison of cohort characteristics and outcomes between FirstPLUS and OxPLUS cohorts, including all cases with evaluable 3D placental scan data.....	55
Table 7: Subgroup analysis of OxPLUS cases comparing cohort characteristics and outcomes between the evaluable cohort and cases of automated quality control failure.....	56
Table 8: Characteristics of FirstPLUS subgroup for centile chart construction	69
Table 9: Akaike information criterion assessment of quantile regression models for prediction of each OxNNet-derived parameter from gestational age in days.....	74
Table 10: Akaike information criterion and GAMLSS family for each Lambda-Mu-Sigma derived model	75
Table 11: Percentage of cases at various centiles observed in study cohort (Obs) vs predicted by final model (Pred) for A) placental volume, svFMBV1 and svFMBV2; and B) svFMBV3, svFMBV4, and svFMBV5	78
Table 12: Westgard multi-control rules.....	89
Table 13: Conditional logic and resulting textual outputs for self-assessment feedback tool for placental volume and svFMBV2	94
Table 14: Multi-rule approach to quality control	97
Table 15: Values of p_0 and h used for CUSUM analysis.....	98
Table 16: Summary of learning events required to achieve competence in first trimester 3D placental ultrasound with power Doppler.	101
Table 17: Sonographer performance as rated using the OxNNet feedback toolkit before and after receiving feedback.....	104

Table 18: Summary of the incidence of multi-rule quality control failures across subgroup cohorts 1–3	106
Table 19: Decisions regarding degrees of freedom applied using restricted cubic splines to each continuous candidate variable	125
Table 20: Recommendations for variables to include in predictive models based on ANOVA and backwards selection	127
Table 21: Final components of multivariable models for the prediction of fetal growth restriction and pre-eclampsia	128
Table 22: Area under the ROC curve for each model including apparent and optimism corrected AUC from the training dataset, and AUC derived from external validation	130
Table 23: Sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) for each model on external validation calculated for false positive rates (FPR) of 10%, 30%, and 50%..	131
Table 24: Pairwise comparison of AUC between models for prediction of fetal growth restriction and pre-eclampsia	133
Table 25: Performance of each model in subgroups for external validation	135
Table 26: Summary of key characteristics and outcomes for the combined OxPLUS and FirstPLUS cohort (n = 6246).....	157
Table 27: Summary of missing data for each novel marker (n = 6246)	158
Table 28: Summary of demographic characteristics for intra-operator and inter-operator reproducibility analysis cohorts.....	176
Table 29: Intra-operator agreement analysis	178
Table 30: Inter-operator agreement analysis.....	178

LIST OF ABBREVIATIONS

2

2D: Two-dimensional

3

3D: Three-dimensional

A

ACOG: American College of Obstetricians and Gynecologists

AIC: Akaike information criterion

ANOVA: Analysis of variance

APS: Antiphospholipid syndrome

AUC: Area under the curve

B

BCCGo: Box-Cox Cole and Green distribution (ordinary)

BCE: Before common era

BCPEo: Box-Cox power exponential (ordinary)

BMI: Body mass index

BP: Blood pressure

BW: Birthweight

C

CI: Confidence interval

CPU: Central processing unit

CUSUM: Cumulative sum

D

DI: Dispersion index

DIC: Disseminated intravascular coagulation

E

EFW: Estimated fetal weight

F

FGR: Fetal growth restriction

FI: Flow index

FirstPLUS: First-trimester Placental UltraSound study

FMBV: Fractional moving blood volume

FMF: Fetal Medicine Foundation

FSA: Functional spiral arteries

FTPV: First trimester placental volume

G

GAMLSS: Generalised additive models for location, scale and shape

GDM: Gestational diabetes mellitus

GT: Generalised t distribution

I

IQR: Interquartile range

IRAS: Integrated research application system

ISSHP: International Society for the Study of Hypertension in Pregnancy

ISUOG: International Society of Ultrasound in Obstetrics and Gynecology

IVS: Intervillous space

J

JSU: Johnson's SU distribution

L

LDA: Low dose aspirin

LGA: Large for gestational age

LMS: Lambda-Mu-Sigma

LOGNO: Log normal distribution

M

MoMs: Multiples of the median

MVP: Myometrial vascular plexus

N

NHS: National Health Service

NICE: National Institute for Health and Care Excellence

NIfTI: Neuroimaging informatics technology initiative

NT: Nuchal translucency

O

OR: Odds ratio

OxNNet: Oxford Neural Network

OxPLUS: Oxford Placental UltraSound study

P

PAPP-A: Pregnancy associated plasma protein A

PE: Pre-eclampsia

PI: Pulsatility index

Q

QC: Quality control

QR: Quantile regression

R

RR: Relative risk

S

SEP: Skew exponential power

SGA: Small for gestational age

SHASH: Sinh-arcsinh

SHASHo: Sinh-arcsinh (ordinary)

SLE: Systemic lupus erythematosus

SMFM: Society for Maternal-Fetal Medicine

SNG: Sub-noise gain

STROBE: Strengthening the reporting of observational studies in epidemiology

svFMBV: Single vessel fractional moving blood volume

SVR: Surface-to-volume ratio

U

UPI: Uteroplacental interface

UtA: Uterine artery/arteries

V

VFI: Vascularisation flow index

VI: Vascularisation index

VOCAL™: Virtual Organ Computer-aided AnaLysis

W

WHO: World Health Organization

[1] INTRODUCTION

The importance of the placenta for fetal nourishment, development and survival has been understood since at least the 5th century BCE, the time of the Ancient Greeks.¹ More than two thousand years later, we now understand that dysfunction of the placenta, and of maternal-placental interaction, contributes to a spectrum of interrelated syndromes including fetal growth restriction (FGR) and pre-eclampsia (PE).^{2,3} The global impact of these ‘placental insufficiency’ syndromes on maternal and perinatal morbidity and mortality is substantial.⁴⁻⁷

While there is considerable scientific research on how the human placenta works and what can go wrong, important questions remain — especially about the basic vascular causes of FGR and PE.

Furthermore, our capacity to accurately identify individuals at risk of placental insufficiency remains limited, which poses a significant challenge to the development and implementation of effective risk mitigation strategies. Novel placental biomarkers, particularly those that reflect the behaviour of the uteroplacental vasculature, are consequently of considerable interest.

The high-level aim of this work is therefore to better understand uteroplacental vascular changes linked to FGR and PE, and to improve clinical prediction of these conditions.

1.1 Clinical motivation

1.1.1 Fetal growth restriction

It is proposed that all fetuses have an inherent ‘growth potential’ — the size to which they will grow under optimal conditions. FGR is the failure of a fetus to achieve its growth potential, frequently presenting as diminished growth velocity or an estimated fetal weight (EFW) or birthweight (BW) that is lower than expected for the given gestational age. Criteria for the antenatal diagnosis of FGR have been agreed by Delphi consensus (summarised in **Table 1**, page 20), but there is a lack of agreement on how to ascertain growth potential itself.⁸ Consequently, small for gestational age (SGA), generally defined as an EFW or BW <10th centile, is frequently used as a proxy for FGR in medical research and in clinical practice. This is troubling because some babies that are SGA are merely constitutionally small (that is, they are achieving their growth potential) without other evidence of FGR. Similarly, a fetus or neonate whose growth has faltered may not fall below the 10th centile but may well be failing to achieve their growth potential and could therefore be at risk of adverse outcomes. This matters because failure to recognise and appropriately manage a growth restricted fetus risks significant harm.⁹ But conversely, obstetric intervention in cases of SGA without FGR also has the potential to harm, particularly if this results in iatrogenic preterm birth. Indeed, national guidance in the UK highlights the importance of avoiding preterm or early term birth, except where there is sufficient expectation that doing so will reduce the risk of serious adverse events.¹⁰

How we define FGR in both clinical practice and research settings is therefore extremely important, and definitions that rely solely on absolute thresholds of EFW or BW may not be applicable to all cases of FGR. We cannot ignore the wealth of literature that exists in relation to SGA, which remains an important (albeit limited) marker for FGR, but cautious interpretation is required with respect to the implications for true FGR.

Furthermore, uteroplacental insufficiency is not the sole cause of FGR. Associations can also be observed with FGR and maternal medical conditions¹¹⁻¹³, drugs (both prescribed and recreational)^{14,15}, poor nutrition¹⁶, infection,¹⁷ chromosomal abnormalities¹⁸⁻²⁰, fetal cardiac anomalies^{21,22}, gastroschisis²³, and neural tube defects.²⁴ However, uteroplacental insufficiency is the most common cause of FGR and is the primary focus of this thesis.²⁵

Table 1: Delphi consensus diagnostic criteria for fetal growth restriction

Early FGR	Late FGR
Gestational age <32 weeks	Gestational age ≥32 weeks
Any one of: - Abdominal circumference <3rd centile - Estimated fetal weight <3rd centile - Umbilical artery absent end diastolic flow Or: - Abdominal circumference or EFW <10th centile combined with... - Uterine artery pulsatility index >95th centile and/or umbilical artery pulsatility index >95th centile	Any one of: - Abdominal circumference <3rd centile - Estimated fetal weight <3rd centile Or at least two out of three of the following: - Abdominal circumference or EFW <10th centile - Abdominal circumference or EFW crossing >2 quartiles on non-customised growth charts - Cerebro-placental ratio <5th centile or umbilical artery pulsatility index >95th centile

1.1.2 Pre-eclampsia

PE complicates approximately 2–5% of pregnancies^{26,27}, although there are also populations where the incidence is much higher.²⁸ The classical diagnostic criteria for PE are new onset hypertension in pregnancy (generally $\geq 140/90$ mmHg at ≥ 20 weeks' gestation) with proteinuria (typically ≥ 30 mg/mmol if formally assessed) representing renal dysfunction. However, pre-eclampsia is a highly heterogeneous condition, and more recent diagnostic frameworks account for evidence of other systemic dysfunction as an alternative to proteinuria. There are several key diagnostic guidelines which overlap significantly and a comparison between three of the most commonly used is presented in **Table 2** (page 22).²⁹⁻³¹ Most of the differences in recommendations are subtle, but one critical distinction is the exclusion of uteroplacental dysfunction from the American College of Obstetricians and Gynecologists (ACOG) recommendation²⁹, whereas it is included in those of the International Society for the Study of Hypertension in Pregnancy (ISSHP)³⁰ and the National Institute for Health and Care Excellence (NICE).³¹ This difference is important in medical research, because including FGR as part of the diagnostic criteria for PE makes it virtually impossible to distinguish between these two phenotypes of placental insufficiency based on diagnostic codes alone.

Table 2: Comparison of three key diagnostic frameworks for pre-eclampsia

Diagnostic parameter	Diagnostic framework		
	ACOG (2020) ²⁹	ISSHP (2021) ³⁰	NICE (2019) ³¹
General qualifier	≥20/40 gestation with hypertension plus at least one other feature.		
Hypertension	≥140/90	≥140/90	≥140/90
Hypertension qualifier	Measured on two occasions at least 4 hours apart.	At least 2 readings should be averaged to reflect BP for the visit.	Not specified
Proteinuria	≥300 mg per 24-hours, protein/creatinine ratio ≥0.3 mg/mg, or dipstick reading of ≥2+	Protein/creatinine ratio ≥0.3 mg/mg	Protein/creatinine ratio ≥30 mg/mmol, albumin/creatinine ratio of ≥8 mg/mmol, or dipstick reading ≥1+
Haematological complications	Platelet count <100 x 10 ⁹ /L	Platelet count <150 x 10 ⁹ /L, DIC, or haemolysis	Platelet count <150 x 10 ⁹ /L, DIC, or haemolysis
Renal impairment	Serum creatinine ≥1.1 mg/dL or doubling in the absence of other renal disease	Serum creatinine ≥ 90 µmol/L or 1 mg/dL	Serum creatinine ≥ 90 µmol/L or 1.02 mg/dL
Hepatic impairment	Transaminases twice normal concentration	Transaminases >40 IU/L	Transaminases >40 IU/L
Pulmonary oedema	Included	Not included	Not included
Neurological complications	Severe headache	Not included	Eclampsia, altered mental state, blindness, stroke, severe headache, persistent visual scotomata
Uteroplacental dysfunction	Not included	Placental abruption, angiogenic imbalance, FGR, abnormal umbilical artery Doppler waveform analysis, or intrauterine fetal death	FGR, abnormal umbilical artery waveform, stillbirth

1.1.3 Clinical outcomes

Both FGR and PE are worthy of attention as they are associated with a significantly increased risk of adverse pregnancy outcomes.

PE is a major contributor to global maternal mortality, with hypertensive complications accounting for 16% of maternal deaths worldwide (equating to approximately 46,000 deaths a year).⁷ For survivors of pre-eclampsia, there is an association with adverse health outcomes in the longer term.³² Several studies have reported increased risk of cardiovascular disease and stroke in women with pregnancies complicated by PE³³⁻³⁵, greater risk of chronic hypertension^{33,35,36}, and a greater probability of late-onset diabetes.³⁶ PE also represents a disease burden for the fetus and neonate — globally, pre-eclampsia accounts for approximately 26% of perinatal mortality³⁷ and is an important cause of iatrogenic preterm birth.³⁸

1.2 Pathophysiology

The myometrial vascular plexus (MVP) of the placental bed comprises an intricate network of vessels. Originating from the uterine arteries (UtA), the arcuate arteries branch into radial arteries, which give rise to basal arteries and spiral arteries (**Figure 1**).

During early pregnancy, endovascular trophoblasts invade the uterine spiral arteries and, in a process known as placentation, the spiral arteries undergo structural remodelling (probably between 8–18 weeks' gestation³⁹), becoming dilated and inelastic vessels that function independently of maternal vasomotor regulation.⁴⁰ This can be observed in histological studies which reveal changes in spiral artery morphology in pregnancy.^{41,42} Spiral artery remodelling is often thought to occur in two distinct waves with the primary (decidual) phase being completed by around 10 weeks' gestation, and a secondary (myometrial) phase around 4–5 weeks later, completing by 16 weeks' gestation.^{43,44} However, this model is challenged by findings showing a gradual rise in the number of myometrial vessels with endovascular trophoblast at 10–12 weeks' gestation, which supports remodelling as a continually progressive process.^{45–48} There is therefore uncertainty about exactly how this remodelling progresses in healthy pregnancies.

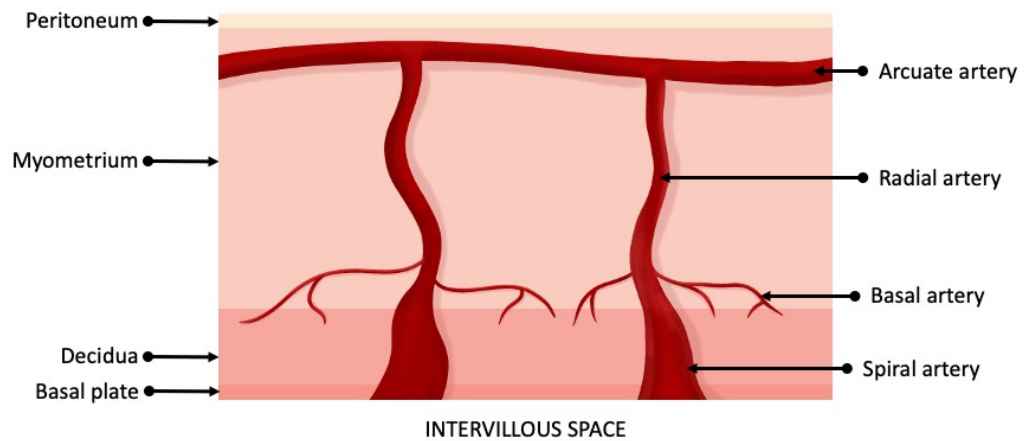
Nevertheless, deficient spiral artery remodelling is well-known to be associated with FGR and PE and is thought to be key to the aetiology of both conditions.^{49–53} The mechanism underlying inadequate remodelling remains incompletely understood but likely involves interactions between the maternal immune system within the decidua and paternally derived antigens presented on trophoblast cells.⁵⁴

The reason why an apparently similar antecedent event leads to one disease phenotype or the other, or manifests as both FGR and PE, is not completely understood and might involve differences in the extent or type of vascular compromise, or differences in the maternal response to this event.^{51,55} Proposed mechanisms to link these histopathological findings to disease outcomes include the occlusion of spiral and basal arteries with atheromatous plaques contributing to infarction^{56,57}, oxidative stress due to ischaemia and reperfusion^{54,58}, and damage to microvilli caused by high pressure 'jets' of blood exiting from poorly remodelled vessels with the corresponding release of syncytial sprouts and trophoblast fragments into maternal circulation contributing to a systemic inflammatory response (the so-called 'jet-hose' theory).^{53,59} Certainly with respect to PE, the 'two-stage' model is generally accepted, whereby the first stage of poor placentation occurs early in

pregnancy prior to clinical manifestation of PE, and the second stage (involving the clinical manifestation of the disease) occurs as a result of factors released from the oxidatively stressed placenta, ultimately contributing to a widespread inflammatory response, endothelial damage, and an increase in platelet activation, thrombosis, and placental infarction.^{3,54,55}

Figure 1: *The vascular anatomy of the myometrial vascular plexus in the pregnant uterus*

Redrawn and adapted from Robertson WB, Uteroplacental vasculature, Journal of Clinical Pathology, 1976.⁶⁰
Not to scale.



1.2.1 Uteroplacental haemodynamics and unanswered questions

Properties of the uteroplacental vasculature can be evaluated *in vivo*. Ultrasonographic measurement of the spiral artery jets as they emerge into the intervillous space (IVS) reveals high resistance changes associated with SGA pregnancies, consistent with the histological changes in spiral artery morphology observed in FGR pregnancies, which adds weight to the 'jet-hose' theory.^{61,62} This is further supported by computational modelling, which suggests that larger jets extending through more than half of the placental thickness (so-called 'mega-jets') can only occur in regions with sparse villous structures, supporting the notion that there is a relationship between blood flow and villous tree development.⁶³

However, it is measurement of upstream blood flow at the level of the uterine arteries that has been adopted into clinical practice. In pregnancy, the uterine arteries generally demonstrate low resistance flow, but high resistance waveform patterns (which can be measured as early as the first trimester) are associated with increased risk of both FGR and PE.⁶⁴⁻⁶⁶ It has been suggested that this occurs as a result of resistance to flow through the spiral arteries, due to failed remodelling, which is transmitted upstream to the uterine arteries where it can be readily assessed sonographically.⁴⁰ Unfortunately, there are several reasons to doubt this explanation, and it is likely that the reality is more complex.

Schaaps *et al.* (2005) have demonstrated that uterine artery vascular resistance in the first 48 hours after delivery is low, similar to the values observed during pregnancy, and that resistance increases gradually over the first seven days of the postnatal period.⁶⁷ This is counterintuitive because myometrial contraction following delivery of the placenta (the 'physiological tourniquet'), would be expected to result in immediate downstream resistance which ought to be transmitted upstream to the uterine arteries and be immediately measurable. Therefore, it is logical that shunting of blood must occur within the MVP somewhere between the uterine arteries and the spiral arteries. Indeed, evidence of these shunts (arteriovenous anastomoses) can be observed in studies of casts of the uterine vasculature.^{67,68} This suggests that in cases demonstrating increased resistance in the uterine arteries, the arteriovenous shunts are likely not to be functioning as normal or that the system has exceeded its capacity to buffer increased downstream resistance. This might explain why uterine artery Doppler abnormalities show better sensitivity and specificity for the more severe cases of PE and FGR. Another possible explanation is that there are haemodynamic changes occurring in the arcuate or radial arteries which influence upstream resistance and could also impact on buffering

capacity. The arcuate and radial vessels are difficult to identify and measure sonographically *in vivo*, but computer modelling suggests that failed spiral artery remodelling alone is unlikely to account for high resistance uterine artery waveforms, and that changes throughout the entire downstream system (particularly the radial arteries) would be necessary to explain the observed upstream resistance.⁶⁹

Additionally, there are instances where changes in the uterine artery waveform occur without any possibility of trophoblast mediated vessel remodelling. This includes haemodynamic changes observed at different phases of the menstrual cycle⁷⁰⁻⁷², with the use of certain drugs and contraceptives⁷³⁻⁷⁶, and even an extremely rare case of third-trimester abdominal pregnancy where the placenta developed entirely outside the uterus.⁷⁷ All of these observations suggest that some regulatory mechanisms operate independently of trophoblast invasion and spiral artery remodelling.

Novel imaging techniques with the capacity to interrogate the deep myometrial vasculature *in vivo* are therefore of interest and have the potential to enhance our understanding of the pathophysiology of FGR and PE as well as providing novel screening targets for these important conditions.

1.3 Risk mitigation and the importance of screening

Early screening plays a pivotal role in risk reduction strategies for FGR and PE, and this concept features in key national and international guidance.^{10,29-31,78-81} The overarching principle is that screening can be conducted early in pregnancy, enabling subsequent care to be tailored accordingly. This approach is particularly valuable when resource constraints limit the feasibility of comprehensive antenatal surveillance for all patients. Individuals identified as high risk for FGR typically undergo more frequent ultrasound assessments of fetal biometry (growth scans), and early identification of suboptimal fetal growth has been shown to improve outcomes.⁹ Those at increased risk for PE may require heightened monitoring through regular blood pressure measurements and urinalysis, and there may also be an opportunity to alter the course of disease progression with pharmacological prophylaxis.⁸²⁻⁸⁴

1.3.1 Screening approaches

Historically, screening for both PE and FGR has relied on the identification of binary risk factors, such as the risk assessments advocated by NICE for PE³¹ and Saving Babies' Lives for FGR.¹⁰ These 'checkbox-style' risk assessments represent a pragmatic approach because they are quick, inexpensive, require little specialist equipment, and they are therefore accessible and scalable. However, this approach risks oversimplification of the relationship between risk factors and outcomes. An alternative approach to prediction involves the development of multivariable models capable of assigning an individualised risk score to each patient based upon important risk factors from the medical history in addition to numerous important biomarkers. This approach is often favoured because it can account for more complex non-linear relationships between predictors and outcomes, it can allow for interaction between predictors, and it can offer a weighting to the assigned level of risk. Perhaps the best known of these is the 'Fetal Medicine Foundation algorithm', a competing risks model which was developed for the SPREE study and has since been refined.⁸⁵⁻⁸⁷ Similar models are available for the prediction of SGA and FGR, but these have not been widely adopted for clinical use.⁸⁸⁻⁹⁴

1.3.2 Prophylaxis and prevention of harm

There is now a wealth of evidence (including several meta-analyses) supporting the use of low-dose aspirin (LDA) in early pregnancy to prevent PE, particularly early-onset and severe PE.⁸²⁻⁸⁴ There is less certainty about the benefit with respect to term PE and FGR. While some studies suggest no impact on the incidence of term PE^{64,84}, there is also evidence to the contrary.^{83,95} Most convincingly, a 2017 meta-analysis including 45 randomised controlled trials has shown a reduction in PE at all gestations (RR 0.57; 95% CI, 0.43–0.75), severe PE (RR 0.47; 95% CI, 0.26–0.83), and FGR (RR 0.56; 95% CI, 0.44–0.70).⁹⁵ Importantly, the prophylactic benefit of aspirin is contingent upon initiation early in gestation, and does not appear to confer benefit if commenced later than 16 weeks' gestation.^{83,95}

While aspirin's preventive role in FGR is less established than for PE, early risk assessment remains valuable for antenatal care planning and targeted growth scans. Assigning ultrasound resources to high-risk patients increases the detection of faltering fetal growth, and missing such cases before delivery significantly raises the risk of adverse outcomes (OR 4.1; 95% CI, 2.5–6.8) compared to when SGA is identified prior to birth.⁹

Thus, enhancing first trimester screening for placental insufficiency could potentially lead to better outcomes for FGR and PE, and exploring new biomarkers that might improve these screening protocols is a worthwhile endeavour.

1.4 First trimester placental biomarkers

Fetal ultrasonography is routinely employed for pregnancy dating, screening for aneuploidy and structural anomalies, as well as ongoing assessment of fetal growth and wellbeing. Ultrasound use during pregnancy is widely regarded as safe, cost-effective, and generally well tolerated by expectant mothers. Consequently, sonographic biomarkers are an attractive screening target, but routine antenatal care often pays limited sonographic attention to the placenta despite its clear visibility on ultrasound from early gestation. Considering the significant role of placental pathology in the aetiology of both pre-eclampsia (PE) and fetal growth restriction (FGR), it is worth exploring whether the placenta itself could provide valuable sonographic biomarkers to enhance the prediction of these conditions.

1.4.1 First trimester placental volume

It has long been established that placental weight at delivery is positively correlated with birth weight.⁹⁶⁻⁹⁸ Clinicians have also been able to assess placental size during early gestation using two-dimensional (2D) ultrasonography since the 1970s^{99,100} and, as early as 1981, it was proposed that rudimentary measurements of placental size might help predict fetal growth restriction.¹⁰¹ The introduction of three-dimensional (3D) ultrasound in the 1990s significantly improved the accuracy of placental measurements, particularly placental volume, and generated renewed interest in early placental assessment for the prediction of adverse outcomes. First trimester placental volume (FTPV) is associated with subsequent SGA, and this has been shown to be independent of PAPP-A and first trimester uterine artery Doppler indices¹⁰², although abnormalities in second trimester uterine artery Doppler velocimetry are also associated with low FTPV.¹⁰³ Prediction of SGA (and LGA) is improved with the addition of FTPV compared with maternal characteristics alone¹⁰⁴, and associations with both SGA and PE are observed to be similar in magnitude to the association with uterine artery impedance at 22 weeks.¹⁰⁵

1.4.1.1 The challenge of placental segmentation

Despite its advantages, segmentation of the placenta using 3D ultrasound remains technically challenging in early pregnancy. Differentiating the placenta from the myometrium can be imprecise, with the level of difficulty influenced by placental location (posterior placentas may be obscured by artefacts from the overlying fetus), attenuation of the ultrasound beam (often due to maternal adiposity), and the presence of uterine lesions (such as uterine leiomyomas).

Manual delineation of the placenta in every two-dimensional plane within a three-dimensional image is regarded as the 'gold standard'. However, this process requires significant expertise and is highly time-consuming, rendering it impractical and difficult to scale, and even in expert hands it can be difficult to identify the true uteroplacental boundary.¹⁰⁶

1.4.1.2 Placental segmentation using VOCAL™

Much of the literature on first trimester placental volume utilises Virtual Organ Computer-aided AnaLysis (VOCAL™), a proprietary segmentation tool which requires capturing a 3D image of the placenta and generating a series of 2D slices by rotating around a user-defined axis. The number of slices depends on the chosen rotation angle (12°, 15°, 18°, or 30°). In each slice, the placental contour must be manually traced, after which the software interpolates these contours to produce a 3D volume. Because this method relies on geometric interpolation, it may be less accurate when applied to irregularly shaped organs¹⁰⁶ — a notable limitation given that placental morphology in the first trimester is often non-uniform.¹⁰⁷ Despite its commercial availability, VOCAL™ remains time-consuming and requires substantial operator training, making it unsuitable for routine screening. Additionally, it is proprietary software that necessitates manufacturer-specific ultrasound equipment. Concerns have also been raised regarding the reproducibility of volume estimates, with certain software iterations yielding inconsistent results.¹⁰⁸

1.4.1.3 *Semi-automated placental segmentation with machine learning*

Machine learning has been successfully applied to the problem of placental segmentation, with the random walker technique offering a more efficient and less operator-dependent approach to placental segmentation.^{106,109} This method involves labelling a small number of pixels or voxels as either placental tissue or surrounding structures (a process known as ‘seeding’). Specialised software then calculates the probability that each remaining pixel or voxel belongs to the placenta, thereby delineating its borders.¹⁰⁹ Compared to VOCAL™, the random walker algorithm is better suited to irregular placental shapes and demonstrates superior observer reliability.¹⁰⁶ It also requires minimal training and is non-proprietary. Nonetheless, the method is not entirely operator independent, and although the process is semi-automated it remains too time-consuming and laborious for placental volume to become a viable population-wide screening test.

1.4.1.4 *Automated placental segmentation with deep learning*

Deep learning, a specialised subfield of machine learning, has become a cornerstone of modern approaches to automated image analysis. It relies on artificial neural networks, computational models inspired by the structure of biological neurons, to learn complex patterns from data. Rather than being explicitly programmed with rules, these systems improve their performance through exposure to large datasets. In medical imaging, this training data — known as ‘ground truth’ — typically consists of expert-annotated images that serve as reference standards. The richness and accuracy of this ground truth are critical. When trained effectively, deep learning models can deliver high levels of precision and reproducibility, offering scalable solutions that complement or even surpass manual interpretation.¹¹⁰

Using a ground truth dataset of >1000 placental scans segmented with the random walker technique (with 10% validated using manual segmentation), the *Placental Imaging Research Group* (Nuffield Department of Women’s and Reproductive Health, University of Oxford) have developed OxNNet — a fully convolutional neural network capable of entirely automated placental segmentation in a matter of seconds.^{111,112} When OxNNet segmentation is compared with the random walker segmentation algorithm, OxNNet achieves a Dice similarity coefficient of 0.84, indicating strong spatial agreement with ground truth annotations. Moreover, volume estimates derived from OxNNet performed at least

as well as the proprietary VOCAL™ software in predicting SGA, with areas under the receiver operating characteristic curve of 0.65 (95% CI 0.61–0.69) and 0.60 (95% CI 0.55–0.65) respectively.^{106,113}

With OxNNet, large scale studies of the predictive utility of placental volume, and the potential incorporation of derived metrics into population screening protocols, becomes feasible.

1.4.2 Power Doppler and assessment of placental vascularity

The introduction of three-dimensional (3D) Power Doppler imaging has enabled comprehensive evaluation of blood flow across the entire uteroplacental interface. Unlike conventional Doppler modalities, Power Doppler visualises the concentration of moving red blood cells by analysing the amplitude of the Doppler signal, rather than relying on flow velocity or direction (as is the case for spectral Doppler).¹¹⁴ This makes it particularly advantageous for imaging tortuous vessels and detecting low-flow velocities, such as the vessels of the myometrial vascular plexus, as it is largely angle-independent¹¹⁵, and images can be processed to remove low amplitude background noise.¹¹⁶

Power Doppler can be applied in both two-dimensional (2D) and 3D imaging formats to derive indices of vascularity and perfusion. In 3D mode, it offers the potential to capture data from all vessels throughout the MVP. However, its accuracy is influenced by maternal body habitus and machine-specific parameters — especially gain settings¹¹⁴ — which necessitates rigorous standardisation to ensure reliable interpretation.

Importantly, the VOCAL™ tool is capable of generating several indices of 3D vascularity including the vascularisation index (VI) (the number of voxels containing colour), flow index (FI) (a weighted average value of the colour-containing voxels), and vascularisation flow index (VFI) (the product of VI and FI).^{117,118} These metrics have been applied to the assessment of placental vascularity^{55,119} but they are unstandardised, heavily influenced by gain settings^{118,120}, and lack of accounting for individual patient's unique tissue attenuation makes inter-patient comparison unreliable. Therefore, VOCAL™ cannot reliably be used to assess placental vascularity and an alternative approach is required.

1.4.2.1 Fractional moving blood volume

To address the variability introduced by machine-specific parameters and tissue-dependent attenuation, fractional moving blood volume (FMBV) and 3D-FMBV have emerged as promising semi-quantitative indices of tissue vascularity derived from Power Doppler imaging.¹¹⁴ Originally proposed

by Rubin and colleagues, FMBV expresses the amplitude of Doppler scatterers, primarily red blood cells, as a proportion of a locally defined reference signal representing maximal vascularity.¹²¹⁻¹²³ This approach enables relative comparison of perfusion across regions, even when absolute amplitude values are confounded by depth, gain, or beam path heterogeneity.

The reference signal is typically extracted from a region presumed to contain 100% blood volume (and therefore 100% signal amplitude), such as the lumen of a large vessel. Early work on FMBV has extrapolated a reference value from a region of interest using cumulative signal curves, but the accuracy of this method is limited by the size of the vessel, background noise, and the absence of a clear inflection point (or 'knee point') in tissues with high vascularity.¹²⁴

Building on these principles, OxNNet has introduced an automated framework for calculating single-vessel 3D FMBV (3D-svFMBV), which selects a stable reference vessel and normalises perfusion indices from surrounding regions accordingly.¹²⁴ This assumes that beam path artefacts affect both reference and target vessels similarly, thereby preserving the relative integrity of the measurement. OxNNet is capable of calculating 3D-svFMBV at five different sites within the myometrial vascular plexus and placenta, offering spatially resolved insights into regional perfusion (**Table 3**, page 36).

Preliminary data suggest that reduced 3D-svFMBV at the uteroplacental interface and within the intervillous space may be associated with pre-eclampsia (PE), whereas normotensive small-for-gestational-age (SGA) pregnancies do not exhibit the same pattern.¹²⁵ This raises the possibility that 3D-svFMBV could serve as a discriminatory marker between these phenotypes. However, the significance of 3D-svFMBV at other anatomical sites within the uteroplacental unit remains unexplored.

Figure 2: The position of the five OxNNet-derived 3D-svFMBV locations with respect to the placenta and myometrium

A) Schematic representation of the locations with the area of interest in green and the UPI in red (not to scale)
 B) The same locations superimposed on a 2D first trimester ultrasound image

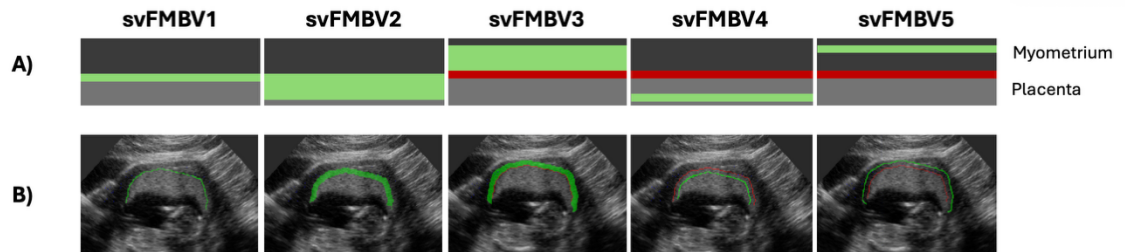


Table 3: Description of the OxNNet-derived 3D-svFMBV locations

Identifier	Location	Target
3D-svFMBV1	1mm thickness from UPI into the placenta	UPI
3D-svFMBV2	5mm thickness from UPI into the placenta	IVS
3D-svFMBV3	5mm thickness from UPI into the myometrium	MVP
3D-svFMBV4	1mm thickness at 5mm distance from UPI into the placenta	Distal IVS
3D-svFMBV5	1mm thickness at 5mm distance from UPI into the myometrium	Proximal MVP

1.5 Research aims and hypotheses

To construct robust, prescriptive centile charts for first trimester placental volume and svFMBV using automated segmentation, and to assess the association of derived z-scores with adverse perinatal outcomes.

To design and implement a structured training programme for sonographers in advanced placental ultrasound, and to develop and evaluate quality assurance tools (including feedback and CUSUM analysis) for monitoring measurement reliability.

To introduce new quantitative markers of placental and vascular morphology, such as uteroplacental interface surface area and spiral artery jet characteristics, and to investigate, on an exploratory basis, their associations with adverse pregnancy outcomes.

To develop multivariable predictive models for FGR and PE incorporating OxNNet-derived imaging biomarkers, maternal history, and serum markers, and to assess their performance using internal validation.

To externally validate the predictive models developed in Chapter 5 using the OxPLUS cohort, and to assess their generalisability.

To assess the intra- and inter-operator reproducibility of OxNNet-derived placental metrics and of the novel markers derived in Chapter 6, and to establish their reliability for research use.

1.6 Publications and presentations arising from this work

1.6.1 Publications in peer reviewed journals

1. **Mathewlynn S**, Starck LN, Yin Y, Soltaninejad M, Swinburne M, Nicolaides KH, Syngelaki A, Contreras AG, Bigiotti S, Woess EM, Gerry S, Collins S. ***Reference charts for first-trimester placental volume derived using OxNNet***. Ultrasound Obstet Gynecol. 2025 Sep;66(3):337-346. doi: 10.1002/uog.29300. Epub 2025 Aug 1. PMID: 40746163; PMCID: PMC12401500.
2. **Mathewlynn S**, Starck LN, Wright D, Yin Y, Soltaninejad M, Nicolaides KH, Syngelaki A, Contreras AG, Bigiotti S, Woess EM, Swinburne M, Collins S. ***First-trimester Placental Ultrasound (FirstPLUS) study: prediction of fetal growth restriction using OxNNet-derived first-trimester placental volume***. Ultrasound Obstet Gynecol. 2026 Jan;67(1):49-59. doi: 10.1002/uog.70146. Epub 2025 Dec 6. PMID: 41351882; PMCID: PMC12757825.
3. **Mathewlynn S**, Starck LN, Yin Y, Soltaninejad M, Swinburne M, Nicolaides KH, Syngelaki A, Contreras AG, Bigiotti S, Woess EM, Gerry S, Collins S. ***Reference charts for first-trimester placental three-dimensional fractional moving blood volume derived using OxNNet***. Ultrasound Obstet Gynecol. 2026 Jan 7. doi: 10.1002/uog.70161. Epub ahead of print. PMID: 41499517.

1.6.2 Oral presentations at international conferences

1. **Volume, vascularity, and artificial intelligence: What can we learn from first trimester 3D placental sonography?** *iMed conference*, Lisbon, October 2023
2. **Artificial intelligence and the first trimester placenta**, *IFPA conference*, Montreal, September 2024
3. **A fully automated tool to screen for fetal growth restriction in the first trimester**, *RCR Global AI conference*, London, February 2025
4. **Fully automated first trimester placental volume for the prediction of fetal growth restriction: Results of the FirstPLUS study**, *BMFMS conference*, Belfast, May 2025

1.6.3 Poster presentations at international conferences

1. Mathewlynn S, Yin Y, Adu-Bredu T, Soltaninejad M, Shipp K, Farmer J, Hedgecott D, Collins S
Reproducibility of first trimester placental volume assessment with fully automated placental segmentation using the OxNNet toolkit *IFPA conference*, Montreal, September 2024

[2] GENERAL METHODS AND DATA SOURCES

2.1 Introduction

The data analysed in this thesis are drawn from two studies, the First Trimester Placental Ultrasound Study (FirstPLUS) and the Oxford Placental Ultrasound Study (OxPLUS), the details of which are discussed in this chapter.

A key goal of this study is to create and externally validate predictive models for placental insufficiency, utilising OxNNet-derived placental metrics. To achieve this, two separate datasets from different clinical environments and populations were generated, supporting comprehensive model training and evaluation across a broad spectrum of patient groups. This methodology not only supports external validation but also enables a targeted assessment of how OxNNet-derived metrics contribute to predictive modelling in placental insufficiency.

The FirstPLUS dataset was collected at a world-renowned centre specialising in fetal medicine and diagnostic ultrasound, making it well-suited for developing both reference charts and predictive models. By contrast, the OxPLUS dataset was gathered in a standard NHS maternity ultrasound department, offering an excellent basis for external validation. Additionally, this setting allows investigation into training and quality control requirements within a context more typical of UK healthcare and provides the chance to conduct repeated measures for analysing the reproducibility of the assessed metrics.

2.1.1 Statement of personal contribution

With the exception of the 3D placental ultrasound data, the FirstPLUS dataset was provided in full by the local study team. The ultrasound data were received in raw format and processed locally by colleagues in the Placental Imaging Research Group, Nuffield Department of Women's and Reproductive Health, University of Oxford. I contributed clinical input to stage 2 image quality control and undertook stage 3 image quality control, as detailed below.

As principal investigator of the OxPLUS study, I have led its set-up and implementation under the supervision of Prof. Sally Collins (chief investigator), including the development of study protocols and standard operating procedures. I designed and delivered a comprehensive training programme for sonographers involved in data acquisition and contributed to the collection of repeated measures data. I have also been solely responsible for all data management and cleaning activities, including extraction of clinical data from hospital electronic patient records, except where otherwise stated.

2.2 Methodology for FirstPLUS and OxPLUS

The methodologies utilised in the FirstPLUS and OxPLUS studies share significant similarities. For conciseness, the following research methods pertain to both studies unless indicated otherwise.

2.2.1 Study design

Both FirstPLUS and OxPLUS are prospective observational cohort studies involving a general population of pregnant individuals undergoing routine first-trimester screening with participants invited to join during their routine ultrasound appointments between 11+2 and 14+1 weeks of gestation. Informed consent was obtained prior to participation. In addition to standard clinical assessments, 3D placental ultrasound imaging was performed during the visit. Participants subsequently continued with routine antenatal care, and pregnancy outcome data were later retrieved from electronic medical records.

Both studies adhered to the principles outlined in the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines.¹²⁶

2.2.1.1 *FirstPLUS study setting*

FirstPLUS was conducted at the Harris Birthright Research Centre for Fetal Medicine, King's College Hospital NHS Foundation Trust, London, UK. Recruitment took place between March and November 2022.

2.2.1.2 *OxPLUS study setting*

OxPLUS was conducted at the John Radcliffe Hospital, Oxford, UK. Recruitment took place between April 2023 and August 2024.

2.2.2 Primary and secondary outcomes

The primary outcome for both studies was FGR, with PE as a secondary outcome.

2.2.3 Ethical approval

The West Midlands – Solihull Research Ethics Committee approved the FirstPLUS study on 8 March 2022 (reference 22/WM/0039, IRAS project ID 308520).

The OxPLUS study was approved by the North West – Greater Manchester Central Research Ethics Committee on 21st February 2023 (reference 23/NW/0035, IRAS project ID 322242).

2.2.4 Eligibility criteria

Eligibility required the following criteria: a singleton pregnancy; maternal age of 18 years or older; attendance for screening between 11+2 and 14+1 weeks' gestation; sufficient understanding of English (spoken or written) or access to appropriate translation services; and, in the judgement of the research team, no significant risk factors or limitations that would impair participation. Exclusion criteria included: detection of multiple gestation (more than one viable fetus) during the scan; non-viable or uncertain viability (absence of fetal heartbeat); identification of major fetal anomalies at the time of scanning or later; confirmed chromosomal abnormalities via prenatal or postnatal testing; or inability to complete the combined screening test (e.g. failure to obtain a nuchal translucency measurement).

2.2.5 Definitions

Fetal growth restriction (FGR) was classified as either birth weight below the 3rd centile, or birth weight below the 10th centile accompanied by abnormal Doppler findings at the final ultrasound prior to delivery. Doppler abnormalities were defined using the reference ranges established by Ciobanu et al.¹²⁷, and included an umbilical artery pulsatility index (PI) exceeding the 95th centile or a cerebroplacental ratio below the 5th centile. Birth-weight centiles were derived from the Fetal Medicine Foundation (FMF) population-based charts.¹²⁸ Hypertensive disorders of pregnancy, including pre-eclampsia (PE) and pregnancy-induced hypertension (PIH), were defined according to the criteria set by the American College of Obstetricians and Gynecologists (ACOG), which are also outlined in **Table 2** (page 22).²⁹ This definition was selected because it does not include features of FGR, enabling the two placental insufficiency phenotypes of PE and FGR to be considered separately.

Gestational diabetes mellitus was diagnosed via oral glucose tolerance testing, following the timing and threshold recommendations of the National Institute for Health and Care Excellence (NICE).¹²⁹ Preterm birth was defined as delivery prior to 37+0 weeks' gestation. Stillbirth referred to birth without signs of life at or beyond 24+0 weeks' gestation, while miscarriage was defined as intrauterine fetal demise confirmed by ultrasound, or birth without signs of life at or before 23+6 weeks' gestation.

2.2.6 Pre-eclampsia risk assessment and aspirin use

The FirstPLUS cohort were risk assessed for PE using the Fetal Medicine Foundation's competing risks algorithm which integrates factors from the maternal medical history with biomarker assessment including mean arterial blood pressure, uterine artery pulsatility index, and serum PAPP-A and PIGF concentration. High risk cases were advised to commence LDA from the date of the scan.⁸⁵⁻⁸⁷

The OxPLUS cohort were not risk assessed using this algorithm, rather they were advised to commence LDA prophylaxis from the date of the scan in accordance with NICE guidance if they had any of the major risk factors or at least two of the moderate risk factors outlined in **Table 4**.³¹ Multifetal pregnancy was an exclusion criterion, so no woman qualified on that basis. In both cohorts the figure reported in Table 6 is therefore the proportion of women advised to take low-dose aspirin, rather than the proportion who took it, since compliance was not recorded. The basis on which that advice was given differed between the studies, being the Fetal Medicine Foundation competing-risks algorithm in FirstPLUS and the NICE criteria in OxPLUS.

The recommended dose of aspirin for both cohorts was 150 mg once a day, to be taken orally in the evening.

Table 4: The major and moderate risk factors for pre-eclampsia as specified in the NICE guideline Hypertension in pregnancy: diagnosis and management (2019)

Major risk factors	Moderate risk factors
• Hypertensive disease in a previous pregnancy • Chronic kidney disease • Autoimmune disease such as systemic lupus erythematosus or antiphospholipid syndrome • Type 1 or type 2 diabetes • Chronic hypertension	• Nulliparity • Age 40 years or older • Pregnancy interval more than 10 years • BMI ≥ 35 at first visit • Family history of PE • Multifetal pregnancy*

*Note that multifetal pregnancy was an exclusion criterion for OxPLUS

2.2.7 Ultrasound scan protocols

2.2.7.1 Ultrasound hardware

Ultrasound examinations were conducted using GE Voluson™ E6 and E8 in the case of FirstPLUS, and E10 systems in the case of OxPLUS (GE Healthcare, Zipf, Austria). All machines were equipped with RAB4-8-D 3D/4D curved-array abdominal transducers operating at 4–8.5 MHz.

2.2.7.2 Ultrasound practitioners

In FirstPLUS, scans were performed by fetal medicine fellows with appropriate certification to perform the routine examinations required for clinical care (including first trimester uterine artery velocimetry), and additional training in first trimester 3D ultrasonography.

In OxPLUS, scans were performed by NHS-employed sonographers who had received additional training in first trimester 3D ultrasonography and first trimester uterine artery velocimetry (the latter of which was not part of routine clinical care at the time of the study). A detailed explanation of the sonographer training protocols is provided in chapter 4.

2.2.7.3 Assessment of gestational age

Pregnancies were dated by crown-rump length (CRL), using the formula determined by Robinson and Fleming (1975)¹³⁰, except for assisted reproduction cases, which used the known date of embryo transfer.

2.2.7.4 The 3D placental scan

To ensure consistency across examinations, machines were pre-configured with standardised settings saved as a dedicated preset (**Table 5**). Optimal probe positioning was first established to enable full visualisation of the placenta in a cross-sectional plane, typically centred on the placental mass. Greyscale gain was adjusted to enhance contrast between the placenta and surrounding tissues. A static greyscale volume was then acquired using the preset parameters, and the resulting image was reviewed to confirm complete placental coverage. If the volume was incomplete or obscured by acoustic shadowing, probe orientation was modified and the acquisition repeated to ensure a full and interpretable placental volume. Power Doppler gain settings were optimised to determine each participant's Sub-Noise Gain setting (SNG). This was achieved by increasing the gain until image saturation ('flooding') occurred, followed by incremental reduction until noise was just eliminated, with no further adjustment (SNG setting).^{120,131} A volume was obtained and evaluated to ensure complete visualisation of both the placenta and underlying myometrium, as well as to assess for the presence of Power Doppler flash artefacts. If the placenta was determined to be incomplete or if more than five Power Doppler flash artefacts were identified, the volume acquisition was repeated.

2.2.7.5 Processing of 3D placental scan images

Ultrasound images were exported directly from the scanning devices as uncompressed 'Kretzfiles' (a proprietary format used by GE ultrasound machines with .vol extension) and were subsequently analysed by the Placental Imaging Research Group at the Nuffield Department of Women's and Reproductive Health, University of Oxford. Images were converted to NIfTI — a widely used medical imaging format that facilitates further analysis and extraction of metadata — using a 'Kretz Converter'¹³². Placental segmentation, volume estimation, and calculation of svFMBV were performed using the OxNNet software.

Table 5: Ultrasound machine settings for 3D first trimester placental scan acquisition

Menu	Parameter	Setting
Main 2D	XBeam Compound receive imaging	Off
	Speckle reduction imaging	3
	Gray map	16
	Tint Map	Off
	Angle	90°
	Time gain compensation	Far Field
Sub 2D	Gray map	16
	Tint	Off
	Line filter	High
	Persistence	6
	Enhance	2
	Line density	Normal
	Reject	20
	Output thermal index	Normal
3D	Volume angle	85°
	Quality	High1
Power Doppler	Quality	Norm
	WMF	Low 1
	PRF	0.9 kHz
Sub power Doppler	PD Map	5
	Frequency	Low
	Flow Res	Mid 2
	L Filter	2
	Artefact	On
	Smooth	Rise 4
	Line Dens	6
	Smooth	Fall 5
	Ensemble	15
	Balance	150

2.2.8 Placental scan quality control procedures

Due to the planned use of the FirstPLUS dataset in model development, a stringent post-hoc quality control process was implemented including manual review of all images and further scrutiny of outliers. In contrast to FirstPLUS, the OxPLUS dataset is intended to represent the ‘real world’ application of 3D first trimester placental ultrasound, and therefore these data were subject only to OxNNet’s internal quality control procedure, described below.

2.2.8.1 *FirstPLUS quality control (round 1)*

Before segmentation, each image was independently reviewed by a trained sonographer to check for placental completeness and technical quality. Images were excluded if placental capture was incomplete, machine settings incorrect, power Doppler artefact exceeded five flashes, or poor quality prevented clear placental delineation.

2.2.8.2 *FirstPLUS quality control (round 2)*

Automated segmentation of the placenta and amniotic fluid was conducted using both current and prototype OxNNet versions. Outputs were compared via Dice similarity coefficients. Cases with unreliable placental boundaries were manually reviewed and excluded for low concordance. Images were flagged if: (1) placental segmentation agreement was <60%; (2) placental segmentation was <75% and amniotic fluid agreement was <40%; or (3) amniotic fluid segmentation agreement was <40%, regardless of placental results.

2.2.8.3 *FirstPLUS quality control (round 3)*

Outliers in placental volume were determined using the interquartile range (IQR) method for each gestational day. At each time point, the first (Q1) and third (Q3) quartiles were calculated; volumes falling below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$ were identified as outliers. These instances underwent manual review by S.M. and M.S. to evaluate image quality and segmentation accuracy. Outliers were excluded if the image quality was considered inadequate for reliable delineation of the placental contour.

2.2.8.4 *OxPLUS quality control*

Sonographer performance and training requirements were monitored during the OxPLUS study, with sonographers identified as outliers provided with additional training. This process is described in detail in Chapter 4 (page 87). An automated process was also developed and incorporated into OxNNet to identify incomplete placental scans (those where the voxels labelled as being part of the placenta are in contact with the boundary of the image).

2.2.9 Statistical methodology

All statistical analyses were conducted using R version 4.3.1 and RStudio version 2024.09.0+375.

2.2.9.1 Target sample size calculation

The target sample size for both studies was determined in advance using the *pmsampsize* package (v1.1.3)¹³³ and Riley et al.'s (2019) methodology.¹³⁴ The aim was to provide sufficient statistical power for the development of multivariable predictive models. Calculations assumed an outcome with 2% incidence, a 20-parameter model, and a variable target C-statistic (area under the ROC curve). These calculations suggest that for a target C-statistic of 0.9 or 0.8 the required minimum sample size would be 2,171 and 2,992, respectively. The recruitment targets were therefore set at 4000 cases for FirstPLUS, and 3500 for OxPLUS. These targets were pragmatic based on the predicted recruitment timescales and an assumed attrition rate of 20%.

2.2.9.2 Cohort and subgroup comparisons

Each study cohort's characteristics were presented as median (interquartile range, IQR) for continuous variables and as counts with percentages for categorical variables. The Wilcoxon rank-sum test was used to compare continuous variables, while the chi-square test was applied to categorical variables. Where any expected cell count was below five, Fisher's exact test was employed instead. Statistical significance was defined as $P < 0.05$.

Key characteristics and outcomes were compared between evaluable cases in the FirstPLUS and OxPLUS cohorts. For the OxPLUS cohort, additional subgroup comparisons were made based on whether cases passed or failed automated quality control, given all other inclusion criteria were met. Subgroup comparison of QC failures was not possible for the FirstPLUS cohort, as demographic and outcome data were not provided in full for excluded cases, including QC failures.

2.3 Results

In total, 4119 cases were enrolled in the FirstPLUS study. After exclusions, a full summary of which is provided in Figure 3, 3479 cases were available for analysis. For OxPLUS, 3519 cases were enrolled, with 2899 remaining after exclusions, 2767 of which passed the automated placental scan quality control procedure (Figure 4, page 53).

Figure 3: FirstPLUS study summary of exclusions and definition of final cohort

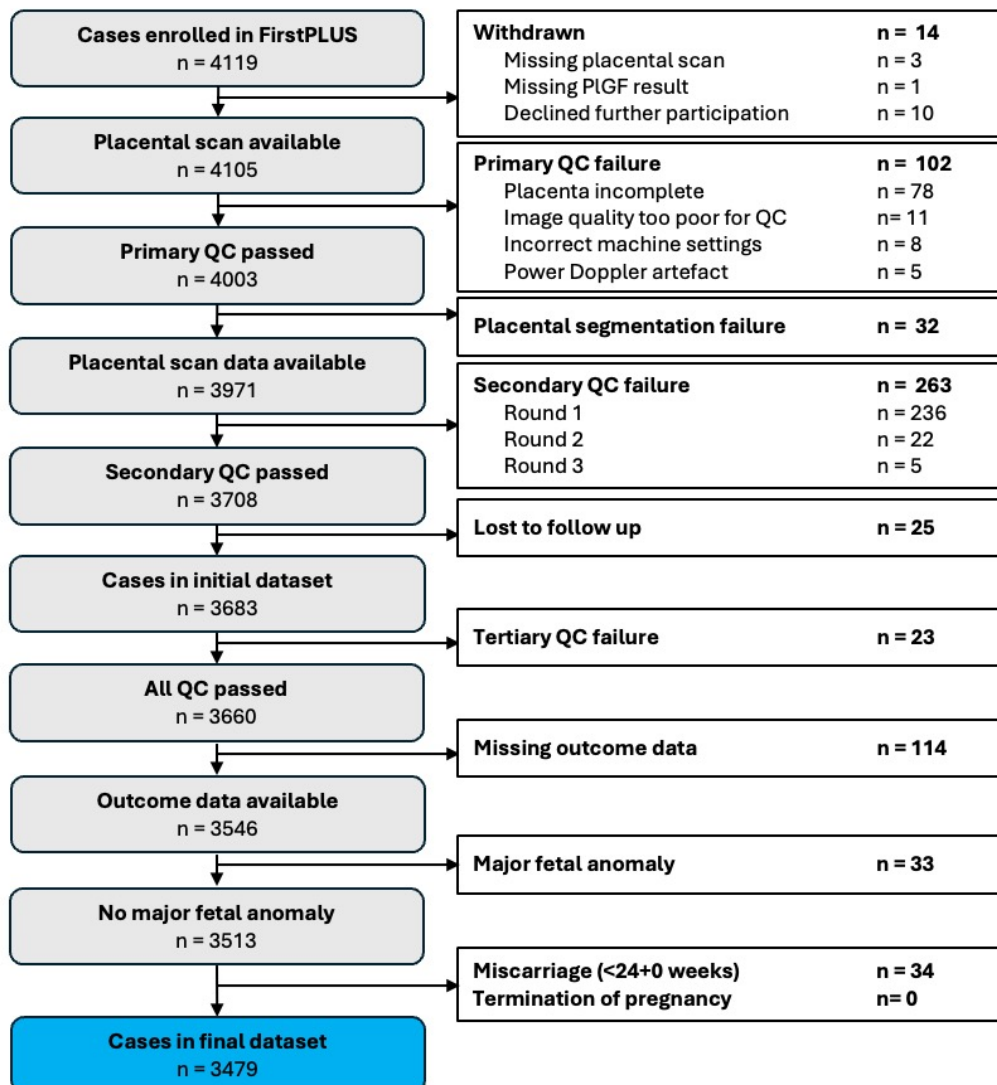
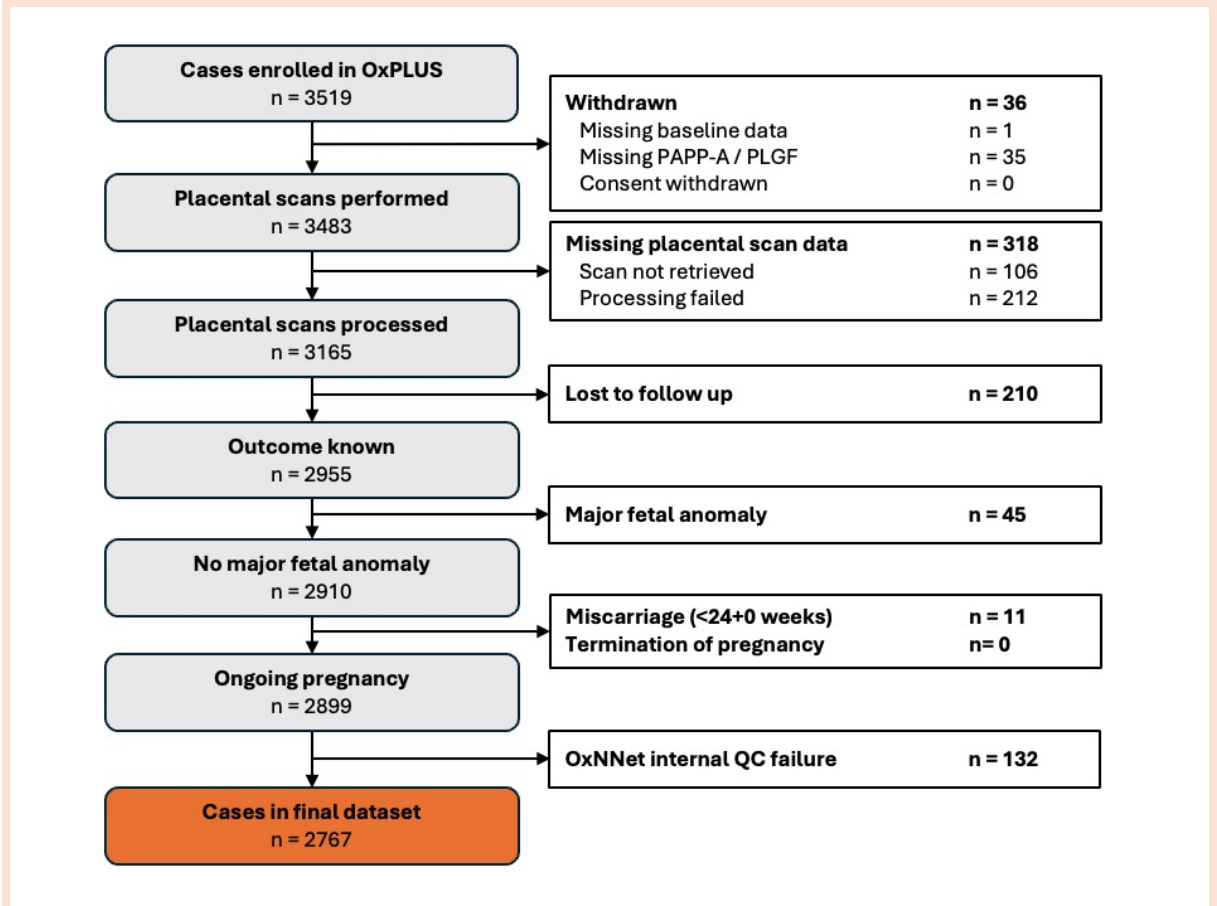


Figure 4: OxPLUS study summary of exclusions and definition of final cohort



2.3.1 FirstPLUS and OxPLUS cohort characteristics

A comparison of characteristics and key outcomes between the FirstPLUS and OxPLUS cohorts (including only those with evaluable placental scans) is shown in **Table 6** (page 55). Statistically significant differences in most characteristics were observed between the two groups, although no difference was seen in parity, chronic hypertension, or autoimmune disease (SLE or APS). For the key outcomes, statistically significant differences were noted in GDM and gestational hypertension. Of particular note, PE was more prevalent in the OxPLUS cohort (4.6% compared with 3.0%, $P < 0.001$) and FGR was less prevalent (4.1% compared with 7.2%, $P < 0.001$). However, the composite outcome of PE or FGR did not differ significantly between the two groups.

A further subgroup comparison of characteristics between those OxPLUS cases with and without evaluable placental scans is shown in **Table 7** (page 56). Cases of QC failure differed significantly with respect to greater maternal age, lower weight, lower BMI, and were less likely to be nulliparous. The core outcomes of interest, including FGR and PE, did not differ significantly between the groups. Some of these differences, while statistically significant, are unlikely to be clinically significant.

Figure 5: Comparison of representation of ethnic categories in the FirstPLUS and OxPLUS cohorts

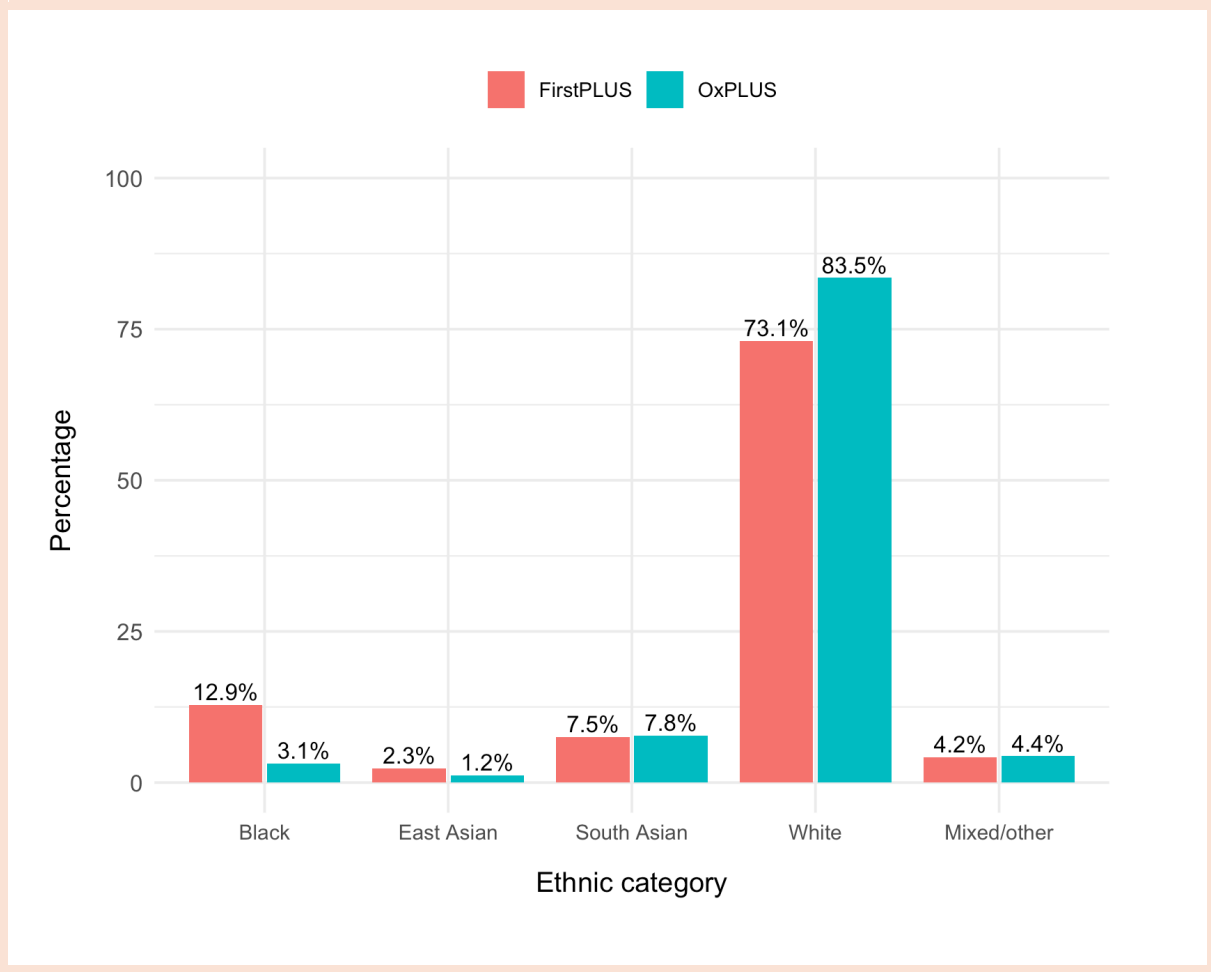


Table 6: Comparison of cohort characteristics and outcomes between FirstPLUS and OxPLUS cohorts, including all cases with evaluable 3D placental scan data

Characteristic	FirstPLUS cohort (N = 3479) Median (IQR) or N (%)	OxPLUS cohort (N = 2767) Median (IQR) or N (%)	P
Maternal age (years)	33.6 (30.4–36.4)	32.4 (29.3–35.5)	<0.001
Maternal weight (kg)	67 (60–77)	69 (61–81)	<0.001
Maternal height (cm)	166 (162–170)	165 (161–169)	<0.001
Body mass index (kg/m ²)	24.2 (21.8–27.8)	25.0 (23.0–29.0)	<0.001
Ethnicity			<0.001
Black	448 (12.9)	87 (3.1)	
East Asian	81 (2.3)	33 (1.2)	
South Asian	262 (7.5)	215 (7.8)	
White	2542 (73.1)	2311 (83.5)	
Mixed/other	146 (4.2)	121 (4.4)	
In vitro fertilisation	263 (7.6)	148 (5.3)	<0.001
Cigarette smoker at presentation	71 (2.0)	115 (4.2)	<0.001
Family history of pre-eclampsia	102 (2.9)	178 (6.4)	<0.001
Nulliparous	1697 (48.8)	1418 (51.2)	0.056
Previous pre-eclampsia	85 (2.4)	94 (3.4)	0.030
Last birth-weight percentile	42.1 (16.7–68.1)	50.0 (21.0–74.0)	<0.001
Interpregnancy interval (years)	2.5 (1.6–4.3)	2.3 (1.5–3.8)	0.002
Chronic hypertension	30 (0.9)	17 (0.6)	0.328
Pre-existing diabetes mellitus			0.002
Type 1	25 (0.7)	10 (0.4)	
Type 2	24 (0.7)	5 (0.2)	
SLE or APS	15 (0.4)	17 (0.6)	0.407
Low-dose aspirin recommended	605 (17.4)	423 (15.3)	0.028
Gestational diabetes mellitus	356 (10.2)	210 (7.6)	<0.001
Gestational hypertension	76 (2.2)	193 (7.0)	<0.001
Pre-eclampsia	103 (3.0)	127 (4.6)	<0.001
Stillbirth	10 (0.3)	8 (0.3)	1.000
Neonatal death	2 (0.1)	3 (0.1)	0.661
Sex of baby			0.969
Female	1728 (49.7)	1369 (49.5)	
Male	1748 (50.2)	1396 (50.5)	
Not known	3 (0.1)	2 (0.1)	
Birth <37+0 weeks' gestation	160 (4.6)	126 (4.6)	0.986
FGR criteria met	249 (7.2)	113 (4.1)	<0.001
Composite outcome (FGR or PE)	327 (9.4)	220 (8.0)	0.060

Table 7: Subgroup analysis of OxPLUS cases comparing cohort characteristics and outcomes between the evaluable cohort and cases of automated quality control failure

Characteristic	QC pass (N = 2767) Median (IQR) or N (%)	QC fail (N = 132) Median (IQR) or N (%)	P
Maternal age (years)	32.4 (29.3–35.5)	33.8 (31.7–36.5)	<0.001
Maternal weight (kg)	69.0 (61.0–81.0)	63.0 (56.0–71.2)	<0.001
Maternal height (cm)	165.0 (161.0–169.0)	165.0 (160.8–169.0)	0.9659
Body mass index (kg/m ²)	25.0 (23.0–29.0)	23.0 (21.0–25.0)	<0.001
Ethnicity			0.1279
Black	87 (3.1)	6 (4.5)	
East Asian	33 (1.2)	5 (3.8)	
South Asian	215 (7.8)	9 (6.8)	
White	2311 (83.5)	108 (81.8)	
Mixed/other	121 (4.4)	4 (3.0)	
<i>In vitro</i> fertilisation	148 (5.3)	6 (4.5)	0.8388
Cigarette smoker at presentation	115 (4.2)	2 (1.5)	0.1723
Family history of pre-eclampsia	178 (6.4)	4 (3.0)	0.1403
Nulliparous	1418 (51.2)	50 (37.9)	0.004
Previous pre-eclampsia	94 (3.4)	3 (2.3)	0.6254
Last birth-weight percentile	50.0 (21.0–74.0)	49.5 (31.0–73.5)	0.4947
Interpregnancy interval (years)	2.3 (1.5–3.8)	2.3 (1.5–3.8)	0.8231
Chronic hypertension	17 (0.6)	1 (0.8)	0.5689
Pre-existing diabetes mellitus			1
Type 1	10 (0.4)	0	
Type 2	5 (0.2)	0	
SLE or APS	17 (0.6)	0	1
Low-dose aspirin recommended	423 (15.3)	10 (7.6)	0.0213
Gestational diabetes mellitus	210 (7.6)	8 (6.1)	0.6299
Gestational hypertension	193 (7.0)	10 (7.6)	0.9286
Pre-eclampsia	127 (4.6)	2 (1.5)	0.1258
Stillbirth	8 (0.3)	0	1
Neonatal death	3 (0.1)	0	1
Sex of baby			0.9284
Female	1369 (49.5)	64 (48.5)	
Male	1396 (50.5)	68 (51.5)	
Not known	2 (0.1)	0	
Birth <37+0 weeks' gestation	126 (4.6)	3 (2.3)	0.2806
FGR criteria met	113 (4.1)	5 (3.8)	1
Composite outcome (FGR or PE)	220 (8.0)	6 (4.5)	0.208

2.4 Discussion and conclusion

This work has resulted in the preparation of two datasets suitable for fitting and external validation of multivariable predictive models for PE and FGR.

2.4.1 Group comparisons

There are statistically significant — but not necessarily clinically significant — differences between the FirstPLUS and OxPLUS cohorts, both with respect to demographic characteristics and core outcomes. For example, the median maternal age of participants in the FirstPLUS study is greater than in OxPLUS (33.6, IQR 30.4–36.4 compared with 32.4, IQR 29.3–35.5, $P < 0.001$) and BMI lower (24.2, IQR 21.8–27.8 compared with 25.0, IQR 23.0–29.0, $P < 0.001$). These differences are small enough that they are unlikely to have a clinically significant impact on the risk of adverse outcome, and it is generally accepted that although the Wilcoxon rank sum test is a robust non-parametric method for detecting differences between groups, in large samples it may yield statistically significant results for very small differences that lack clinical importance.¹³⁵ This is also true for the chi-square test with respect to categorical variables.¹³⁶ Thus, some of the differences observed may simply be statistical artefact.

However, in some cases the observed differences may have greater clinical importance and these are discussed below.

2.4.1.1 Ethnicity

There are differences in ethnicity between the two cohorts ($P < 0.001$), particularly with respect to the representation of black ethnicities — in FirstPLUS 12.9% of participants reported their ethnicity as black compared with only 3.1% of the OxPLUS cohort (**Figure 5**).

The differences in ethnicity are noteworthy both because of the magnitude of difference, and because multiple high-quality studies and meta-analyses have demonstrated that Black women experience higher rates of pre-eclampsia, fetal growth restriction, preterm birth, stillbirth, and neonatal death compared to White women, even after adjustment for maternal characteristics and socioeconomic factors.¹³⁷⁻¹³⁹ Particularly with respect to PE, a systematic review and meta-analysis found that Black women have approximately a two-fold higher risk of both total and preterm pre-eclampsia compared

to White women, even after controlling for maternal characteristics.¹³⁸ A similar-sized effect is seen with respect to FGR.¹³⁹

2.4.1.2 Smoking

There is also a notable difference in the number of smokers between the two cohorts ($P < 0.001$) with only 2.0% of the FirstPLUS cohort being smokers at the time of enrolment compared with 4.2% of the OxPLUS cohort. A dose-dependent association has been demonstrated between maternal smoking and reduced fetal size from early gestation through birth, with the risk and severity of FGR increasing with the number of cigarettes smoked per day.¹⁴⁰ Conversely, smoking is generally considered to be protective with respect to pre-eclampsia, with multiple large epidemiological studies and meta-analyses showing a significantly reduced risk of pre-eclampsia among pregnant smokers compared to non-smokers.^{141,142} A recent meta-analysis reports a relative risk of pre-eclampsia of 0.65 (95% CI 0.58–0.73) for smokers, indicating a substantial inverse association.¹⁴²

2.4.1.3 Pre-eclampsia and fetal growth restriction

Given the differences in smoking and ethnicity, one might expect a higher incidence of PE in the FirstPLUS study, but in fact the opposite is true and PE is more prevalent in the OxPLUS cohort (4.6% compared with 3.0%, $P < 0.001$). This is unlikely to be explained by diagnostic bias, since the clinical criteria for PE diagnosis were the same in both settings, and nor is it likely to be explained by an intervention paradox (i.e. earlier intervention in cases of gestational hypertension preventing PE from developing) because the incidence of gestational hypertension is also substantially lower in the FirstPLUS cohort (2.2% versus 7.0%, $P < 0.001$).

A more likely explanation is the difference in PE risk assessment and aspirin recommendation between the two cohorts, with FirstPLUS following the competing risks multivariable predictive model approach used in the ASPRE trial⁸⁵⁻⁸⁷, and OxPLUS following NICE recommendations.³¹ Tan *et al.* (2018) showed that a multivariable model combining maternal factors and biomarkers (the same used in the FirstPLUS cohort) had substantially higher detection rates for pre-eclampsia, especially preterm pre-eclampsia, than the National Institute for Health and Care Excellence (NICE) guideline approach, at a similar screen-positive rate. For preterm pre-eclampsia, the detection rate was 82.4% with the full multivariable model versus 40.8% with the NICE guideline, both at a 10% screen-positive rate. For all pre-eclampsia, the detection rate was 42.5% with the mini-combined test compared to 30.4% with the

NICE guideline.⁸⁷ The ASPRE trial demonstrated a substantial reduction in preterm PE with aspirin use when risk assessed using the same algorithm (OR 0.38; 95% CI 0.20– 0.74), but no significant reduction was seen in term PE.⁸² This is supported by a meta-analysis of 60 trials that included 36,716 women at elevated risk for pre-eclampsia which showed that aspirin lowers the likelihood of pre-eclampsia in a dose-dependent way (relative risk, 0.82; 95% CI, 0.77 to 0.88), as well as reducing the risk of SGA with a relative risk of 0.84 (95% CI 0.76–0.92).¹⁴³

This could contribute to the lower incidence of PE in the FirstPLUS cohort, despite arguably a higher risk population, but the same effect would be expected for FGR. The greater incidence of FGR might be explained by aspirin having less effect on FGR than on PE, on top of a higher background risk of FGR for the FirstPLUS cohort.

2.4.2 Subgroup comparison

The comparison of QC passes and fails in the OxPLUS cohort suggests that QC failure is associated with lower maternal weight and BMI, rather than higher. This is the opposite of what might be anticipated, since transabdominal ultrasonography is generally more difficult in women of larger body habitus because of signal attenuation and the technical challenges of probe positioning and image optimisation. The explanation is likely to lie in the nature of the automated quality control criterion itself, which does not assess image quality but instead identifies scans in which voxels labelled as placenta are in contact with the boundary of the acquired volume. Failure therefore denotes incomplete capture of the placenta within the field of view rather than a poor quality image. In women of lower weight and BMI the uterus lies closer to the maternal abdominal wall, so that the placenta sits nearer the near-field margin of a fixed acquisition volume and is correspondingly more likely to abut its boundary.

QC failure was also associated with greater maternal age and with parity, cases of QC failure being less likely to be nulliparous (37.9% compared with 51.2%). There is clear evidence from multiple large studies that body mass index (BMI) increases with age among pregnant individuals.^{144,145} Since BMI rises with age, the two associations observed here run in opposite directions, cases of QC failure being older yet of lower BMI. Collinearity with BMI therefore cannot explain the association with age. A more plausible explanation is parity: older women are more likely to be parous, and reduced abdominal wall

tone in parous women would again tend to bring the uterus and placenta closer to the near field of the acquisition volume. Age may therefore be acting largely as a proxy for parity rather than as a direct contributory factor. This interpretation is post hoc and was not formally tested.

It is unfortunate that these findings cannot be corroborated by subgroup analysis of the QC failures in the FirstPLUS cohort, and this must be acknowledged as a limitation.

2.4.3 Conclusion

This work has provided two datasets ready for further analysis. Differences are demonstrated between the study cohorts, both with respect to demographic characteristics and core outcomes. Some difference is expected, since each cohort is drawn from distinct populations, and indeed some difference is advantageous for robust external validation of reference ranges and multivariable predictive models created using one cohort and tested on another.

The observed differences between the two cohorts and the exploration of QC failures highlight the importance and need for deeper exploration of aspirin use, BMI, age, ethnicity and smoking with respect to markers of PE and FGR and prediction of those conditions.

[3] DEVELOPMENT OF CENTILE CHARTS

3.1 Introduction

It is common practice to utilise reference charts when interpreting ultrasound measurements as they allow clinicians to distinguish normal from abnormal findings by comparing individual measurements to population-based centiles. The value of these charts lies in their ability to account for physiological changes across gestation.

Placental size and vascularity evolve with advancing gestational age and this advancement is not necessarily linear. Thus, adjustment for gestational age is critical to avoid misclassification of normal developmental variation as pathology. Z-scores and centile calculations provide standardised metrics for quantifying how far a measurement deviates from the expected mean for a given gestational age, facilitating both clinical decision-making and research comparisons across populations and studies.

There has been much debate about how best to construct centile charts for the evaluation of fetal growth, and much of this discussion is relevant in the context of placental growth and vascular development. Fetal growth charts are constructed using two main approaches: descriptive (reference) charts, which reflect measurements from heterogeneous populations and reflect actual measurements, including those from individuals with varying risk factors (such as the Hadlock chart¹⁴⁶); and prescriptive (standard) charts, which use data from healthy, low-risk populations to define optimal growth trajectories (such as INTERGROWTH-21st and the WHO standards).^{147,148} Descriptive charts may overestimate normality in high-risk groups, and studies comparing these approaches show that prescriptive standards may more accurately identify fetuses at risk for adverse outcomes.¹⁴⁹⁻¹⁵¹

An abundance of centile charts exist for evaluation of fetal growth, covering a range of populations and methodological approaches. For placental volume, however, the choices are extremely limited and the charts that do exist have significant methodological limitations.¹⁵²⁻¹⁵⁴ No such charts exist for uteroplacental FMBV. Consequently, we know very little about what should be considered normal or abnormal with respect to placental volume and vascularity in the first trimester.

3.1.1 Aims

- To develop prescriptive centile charts for placental volume and svFMBV1–5 modelled against gestational age.
- To explore the relationship between the newly derived z-scores and the outcomes of FGR and PE.

3.1.2 Statement of personal contribution

The statistical analysis plans were developed in consultation with Stephen Gerry, Senior Medical Statistician at the Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford. I have prepared all R scripts and conducted all statistical analysis processes independently.

3.2 Methodology

3.2.1 Study population

Participants were excluded if they experienced miscarriage, stillbirth, or pregnancy termination, or if delivery occurred preterm. Exclusions also applied in cases where a major fetal anomaly was detected after the research scan, where there was evidence of fetal growth restriction, or if the neonate required admission to the neonatal unit. Maternal conditions leading to exclusion included diabetes mellitus (either pre-existing or gestational-), hypertensive disorders (chronic hypertension, gestational hypertension, or pre-eclampsia-), antiphospholipid syndrome, and systemic lupus erythematosus. In addition, records lacking complete outcome data were not included in the analysis.

3.2.2 Adequacy of sample size

To evaluate whether the available sample size was adequate for constructing reference charts, we calculated the precision of the data in estimating extreme centiles (specifically the 2.5th and 97.5th percentiles) using the method shown in Equation 1.^{155,156}

Equation 1: Formula to calculate precision in estimation of extreme centiles

$$SE_p = SD \sqrt{\frac{\left(1 + \frac{1}{2} Z_p^2\right)}{n}}$$

In this formula, SE denotes the standard error, SD the standard deviation, Z_p the z-score corresponding to the p^{th} centile, and n the sample size. To achieve a target SE of 0.05 at the 2.5th and 97.5th centiles, a sample size of 1169 was estimated as sufficient, based on a calculated SD of 1 following standardised Box–Cox transformation to approximate a Gaussian distribution for each variable of interest. Using the actual sample size after exclusions ($n = 2547$) and the observed SD, the resulting SE was 0.034, confirming adequacy for centile chart construction. The transformation method was selected using

the *bestNormalize* package¹⁵⁷ in R, and for the purposes of this calculation, variance was assumed to remain constant across gestational age.

3.2.3 Exploration of data

Simple histograms and quantile-quantile plots were constructed to visualise data distributions for each variable of interest. The Shapiro–Wilk test was employed as an objective means of evaluation. Faceted boxplots were constructed for each variable, stratified by gestational age in days, to examine heteroscedasticity.

3.2.4 Models for centile chart construction

Given the evidence of non-normal distributions and potential variation with gestational age, we adopted flexible statistical modelling approaches rather than classical methods based on mean and standard deviation with data transformation.¹⁵⁸ Specifically, quantile regression (QR)¹⁵⁹ and the Lambda-Mu-Sigma (LMS)¹⁶⁰ method — part of the generalised additive models for location, scale and shape (GAMLSS) methodology — were selected for further exploration.

QR analysis was conducted using the *quantreg* package in R.^{159,161} To capture different potential relationships between gestational age and the outcomes several model specifications were explored, including simple linear forms, higher-order polynomials, and flexible spline-based approaches (natural splines and B-splines). Estimation of conditional quantiles was carried out using both the Barrodale–Roberts algorithm (the default linear programming technique) and a penalised regression framework employing the Lasso method.

For LMS modelling, the *lms* function within the *gamlss* package was used.¹⁶² Model fitting involved comparison across a broad set of candidate GAMLSS distributions to identify the most appropriate representation of the data. The distributions assessed included: Box-Cox Cole–Green (BCCGo), Box-Cox Power Exponential (BCPEo), Box-Cox t-Extended (BCTEo), Box-Cox t (BCT), several variants of the Skew Exponential Power (SEP) families 1–4, the Sinh–Arcsinh forms (SHASH, SHASHo), Johnson’s SU (JSU), the Generalised t (GT), Log-Normal (LOGNO), and five types of Skew-t distributions (ST1–ST5). Model fitting was performed with 1000 cycles, a smoothing parameter $k = 2$, and all other options retained at their default settings.

3.2.5 Evaluation of models

The Akaike Information Criterion (AIC) was used to compare models within each modelling approach: to rank the candidate quantile regression specifications against one another, and separately to rank the candidate GAMLSS distributions against one another.¹⁶³ It was not used to compare the two approaches with each other, and the values reported in Table 9 and Table 10 are not on a common scale. The two methods optimise different objective functions: quantile regression minimises an asymmetric check-function loss at a single specified quantile, whereas GAMLSS maximises the full likelihood of a parametric conditional distribution. The difference between their AIC values therefore has no interpretation, and the similarity in their numerical magnitude reflects only the scale of the measurements. The choice between the two approaches was instead made on methodological grounds, and three criteria were applied. The first was whether the method defines a coherent conditional distribution at every gestational age, from which z-scores can be derived directly. Z-scores were required both for the analyses of association with adverse outcome later in this chapter and for the multivariable modelling in Chapter 5, and quantile regression yields them only by interpolating between separately fitted curves. The second was whether the fitted centiles are constrained not to cross, which the LMS approach guarantees by construction, since all centiles derive from a single fitted distribution, and quantile regression does not, since each centile is estimated independently of the others. The third was whether the stronger distributional assumption on which the LMS approach depends is supported by the data, which was assessed by the diagnostic checks described below rather than by AIC. Set against these, quantile regression offers greater robustness, requiring no distributional assumption and allowing each centile to follow the observations in its own region of the distribution. How these considerations were weighed is reported in the results (**3.3.3**, page 73). To further assess performance, centile curves were plotted alongside the observed data, allowing visual inspection of fit across the distribution. In addition, diagnostic checks included histograms and quantile–quantile plots of the derived z-scores to evaluate distributional assumptions, and comparison of the proportion of cases falling below each model-predicted centile with the proportion observed in the sample. All of these assessments were made in the same cases used to fit the models, and therefore describe apparent, or in-sample, fit rather than validated performance.

3.2.6 Exploration of association with adverse outcomes

To assess the potential predictive value of the derived z-scores in the complete FirstPLUS cohort, odds ratios for adverse perinatal outcomes were estimated using univariable binomial logistic regression, with 95% confidence intervals calculated via the Wald method. The outcomes included were FGR, normotensive FGR (FGR without evidence of concurrent hypertension), any hypertensive disorder of pregnancy (including PE and gestational hypertension), PE, and a composite outcome of any FGR or hypertension.

C-statistics with 95% confidence intervals were calculated using the *pROC* package in R.

Forest plots were subsequently generated to visually represent these associations.

3.3 Results

3.3.1 Cohort characteristics

After exclusions, a summary of which is shown in Figure 6, 2547 cases were available for analysis. A summary of cohort characteristics is shown in **Table 8** (page 69).

Figure 6: Summary of exclusions for centile chart development

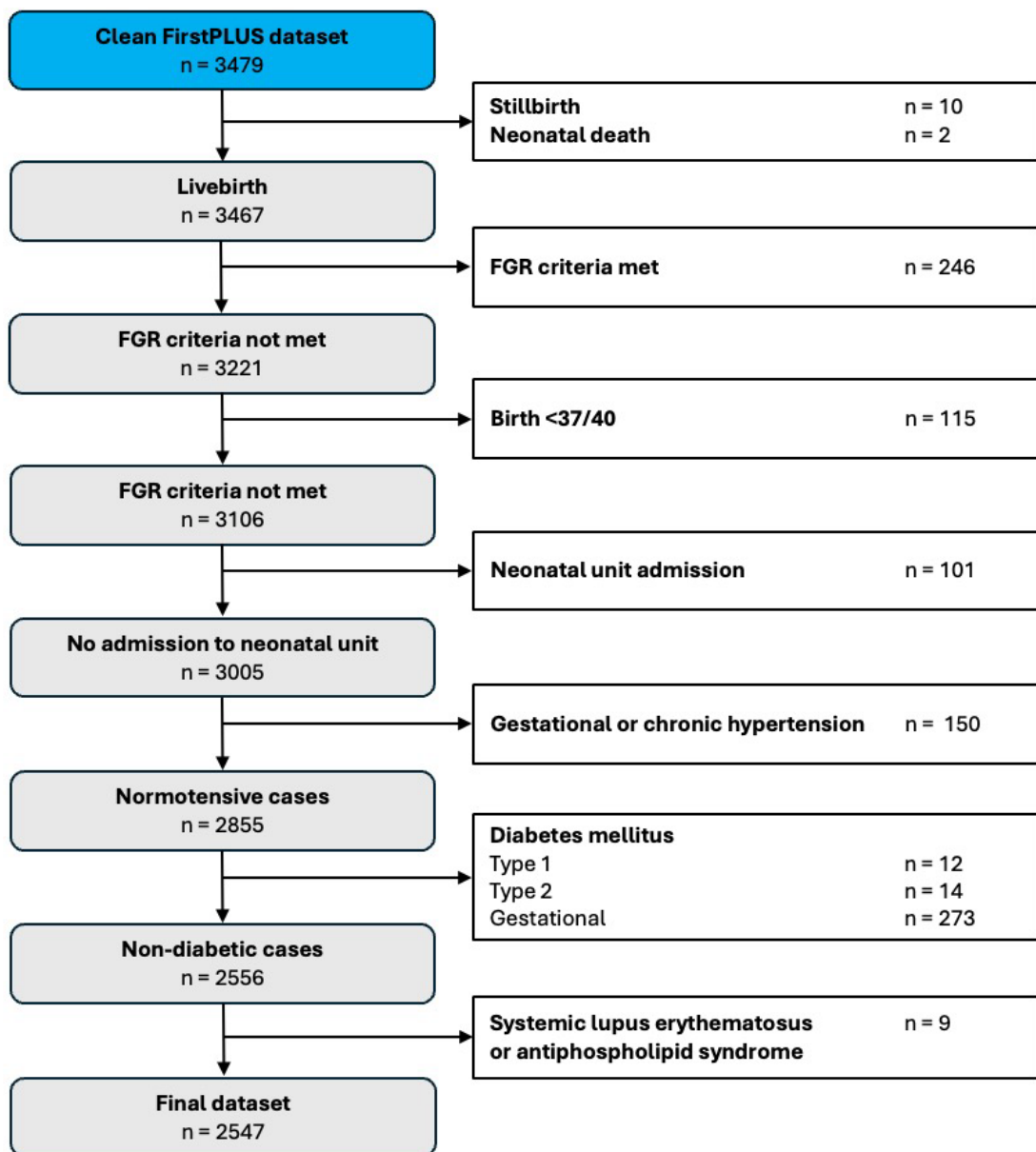


Table 8: Characteristics of FirstPLUS subgroup for centile chart construction

Characteristic	Median (IQR) or N (%)
Maternal age (years)	33.44 (30.40–36.23)
Maternal weight (kg)	66.60 (59.90–75.60)
Maternal height (cm)	166 (162–171)
Body mass index (kg/m ²)	23.90 (21.52–27.21)
Ethnicity	
Black	286 (11.23)
East Asian	52 (2.04)
South Asian	164 (6.44)
White	1944 (76.33)
Mixed/other	101 (3.97)
In vitro fertilisation	165 (6.48)
Cigarette smoker at presentation	46 (1.81)
Family history of pre-eclampsia	70 (2.75)
Nulliparous	1209 (47.47)
Previous pre-eclampsia	40 (1.57)
Last birth-weight percentile	44.39 (19.45–68.62)
Interpregnancy interval (years)	2.43 (1.55–3.96)
Sex of baby	
Female	1264 (49.63)
Male	1283 (50.37)
Not known	0

3.3.2 Analysis of distribution

Simple histograms and quantile-quantile plots are shown in **Figure 7** (page 71), which demonstrates that none of the OxNNet-derived parameters have a normal distribution. This is most pronounced for svFMBV4, which is heavily positively skewed, and svFMBV2 to a lesser extent. The histograms for placental volume, svFMBV1, svFMBV3 and svFMBV5 all appear close to normal, but inspection of the quantile-quantile plots suggest a degree of positive skew. Non-normality of distribution is confirmed by the Shapiro-Wilk test, which generated P values of <0.001 in all cases, strongly suggestive of non-normal distribution.

The faceted boxplots shown in **Figure 8** (page 72) show how the distribution of each parameter varies with gestational age. There is a trend towards increasing placental volume with advancing gestational age, which is not clearly apparent for svFMBV measurements. It is difficult to interpret distribution at the extremes of gestational age due to low case numbers, but nevertheless there are variations in distribution across the gestational age window of interest.

Figure 7: Paired histograms and quantile-quantile plots for each OxNNet-derived parameter

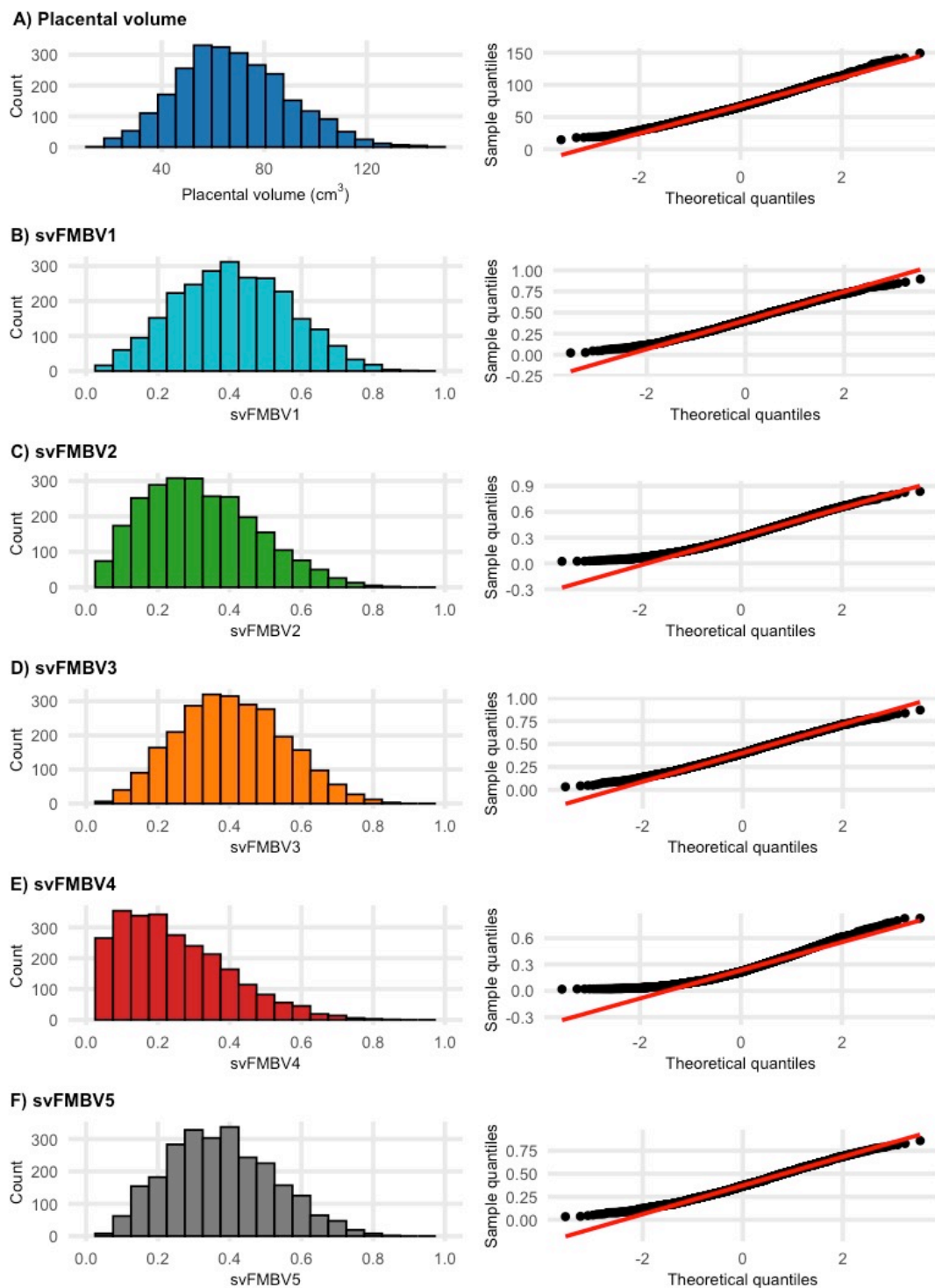
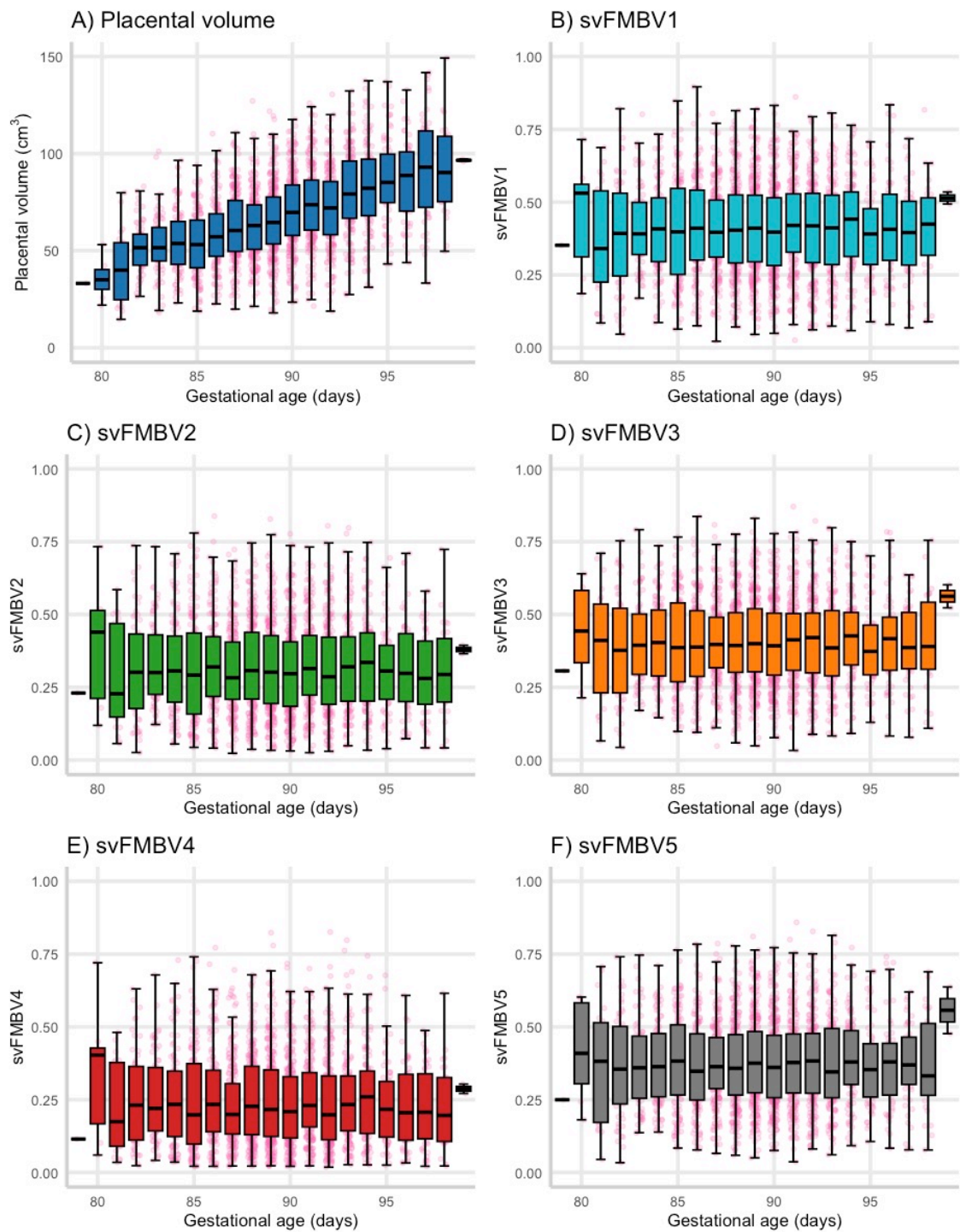


Figure 8: Faceted boxplots for each OxNNet-derived parameter stratified by gestational age in days



3.3.3 Comparison and selection of models

The AIC statistics for the QR models are summarised in **Table 9**. The GAMLSS family selected by the lms toolkit, and the resulting AIC values for LMS models are presented in **Table 10** (page 75). The two sets of AIC values are not comparable with one another, for the reasons given in the methods, and the numerically lower values obtained with the LMS approach were not treated as evidence of superior fit.

Both quantile regression and the LMS method are established approaches to reference chart construction, and each carries its own trade-offs. Quantile regression requires no distributional assumption and estimates each centile from the observations in that region of the distribution, which confers robustness and allows the shape of each centile to follow the data independently. Against this, because the centiles are estimated separately they are not constrained to remain correctly ordered, and the method does not define a conditional distribution from which z-scores can be derived. The LMS method requires the conditional distribution to belong to a single parametric family at every gestational age, which is a stronger assumption, but estimates location, scale and shape as smooth functions of gestational age, so that all centiles derive from one fitted distribution, cannot cross, and yield z-scores directly. For the present application these last two properties were decisive. Z-scores were required both for the analyses of adverse outcome presented below and for the multivariable modelling in Chapter 5, and quantile regression does not provide them without interpolating between separately fitted curves. The parametric assumption on which the LMS approach depends was in turn supported by the fit of the selected distributions, which calibrated closely across the range of target centiles and produced approximately normally distributed z-scores.

The final centile charts for each OxNNet parameter are shown in **Figure 9**.

Table 9: Akaike information criterion assessment of quantile regression models for prediction of each OxNNet-derived parameter from gestational age in days

Modelling approach	Method	Akaike information criterion for each parameter					
		Volume	svFMBV1	svFMBV2	svFMBV3	svFMBV4	svFMBV5
Linear	B&R	22434.66	-1806.84	-1906.37	-2162.46	-2107.87	-2168.95
	Lasso	23121.66	-1807.03	-1907.34	-2161.76	-2109.62	-2168.01
Polynomial 2	B&R	22434.43	-1806.84	-1904.66	-2160.48	-2105.89	-2167.07
	Lasso	22511.89	-1729.21	-1829.40	-2079.70	-2021.81	-2045.41
Polynomial 3	B&R	22436.24	-1805.04	-1902.84	-2160.17	-2105.25	-2166.45
	Lasso	22609.47	-1550.35	-1653.30	-1923.31	-1906.62	-1909.56
Polynomial 4	B&R	22434.15	-1803.08	-1902.63	-2158.22	-2103.56	-2164.46
	Lasso	22744.46	-1366.02	-1444.05	-1767.44	-1703.96	-1748.27
Natural splines	B&R	22437.04	-1803.79	-1902.28	-2158.90	-2103.90	-2164.72
	Lasso	22458.36	-1754.18	-1860.15	-2110.22	-2061.37	-2102.74
B-splines	B&R	22434.78	-1803.21	-1903.00	-2158.17	-2103.67	-2164.45
	Lasso	22481.40	-1761.33	-1882.90	-2111.09	-2091.07	-2121.06

B&R: Barrodale and Roberts method

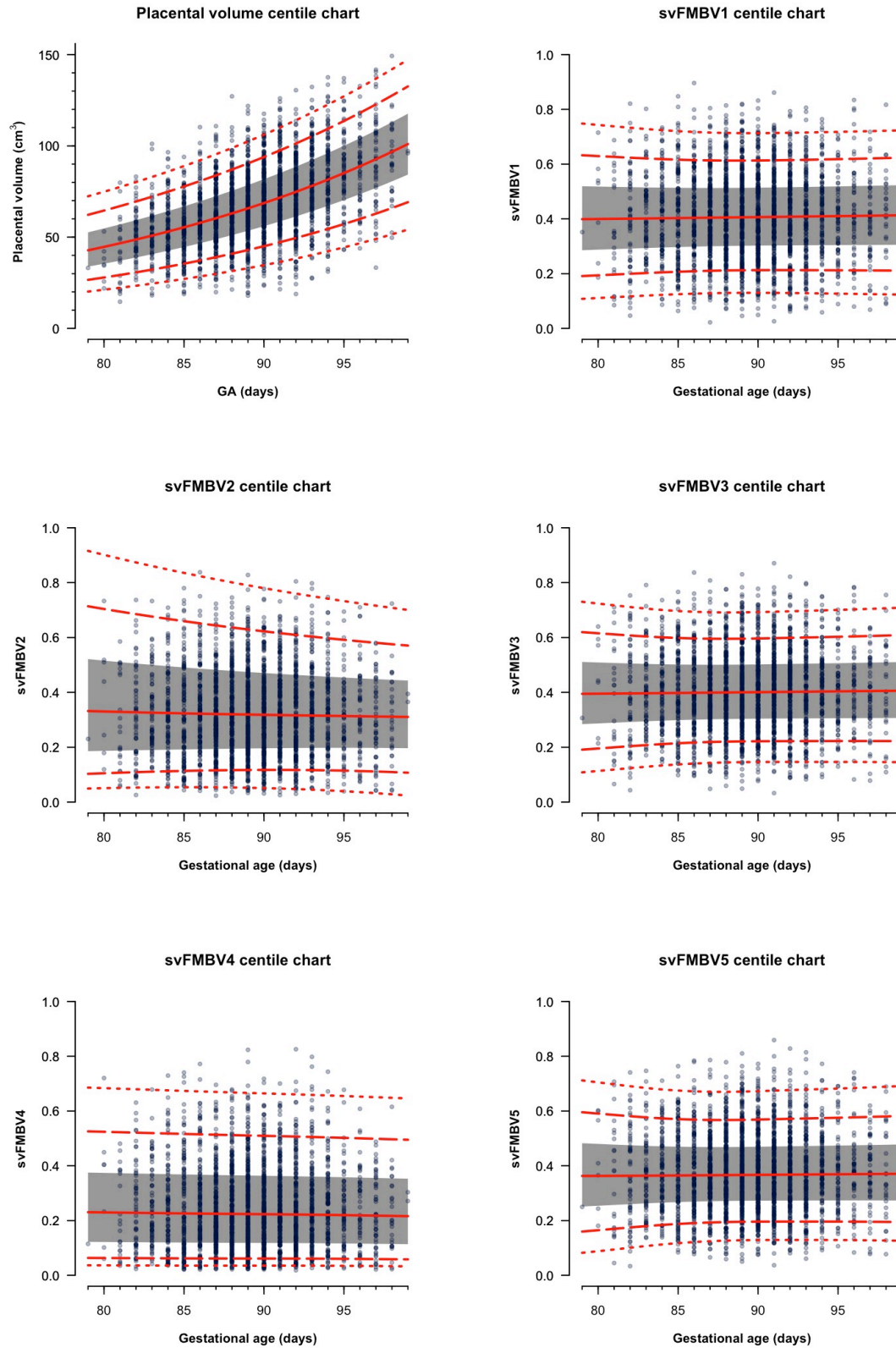
Lasso: Penalised lasso regression

Table 10: Akaike information criterion and GAMLSS family for each Lambda-Mu-Sigma derived model

Parameter	AIC	GAMLSS family
Volume	22135.04	BCCGo
svFMBV1	-2225.73	BCPEo
svFMBV2	-2464.89	SEP3
svFMBV3	-2557.69	BCPEo
svFMBV4	-2956.33	SEP1
svFMBV5	-2612.2	BCPEo

Figure 9: Final centile charts for placental volume and svFMBV1–5.

The solid red line represents the 50th percentile, the grey band the 25th–75th percentile, the dashed red lines the 10th and 90th percentiles, and the dotted red lines the 3rd and 97th percentiles.



3.3.4 Evaluation of centile charts

Table 11 presents a comparison between the predicted and observed centiles across nine target points (1st, 3rd, 10th, 25th, 50th, 75th, 90th, 97th, and 99th) for the six examined models. These comparisons were made in the same 2547 cases used to fit the models, and therefore describe apparent, or in-sample, calibration rather than validated performance; such estimates are optimistic by construction and cannot detect overfitting. On this basis there was close concordance between model-predicted and sample-observed centiles for most parameters and most target points, including those at the distribution extremes. The exception was svFMBV4, for which only 2.55% and 8.32% of cases fell below the predicted 3rd and 10th centiles respectively, indicating that the selected SEP1 distribution places its lower centiles too low and fits that parameter less well in the lower tail.

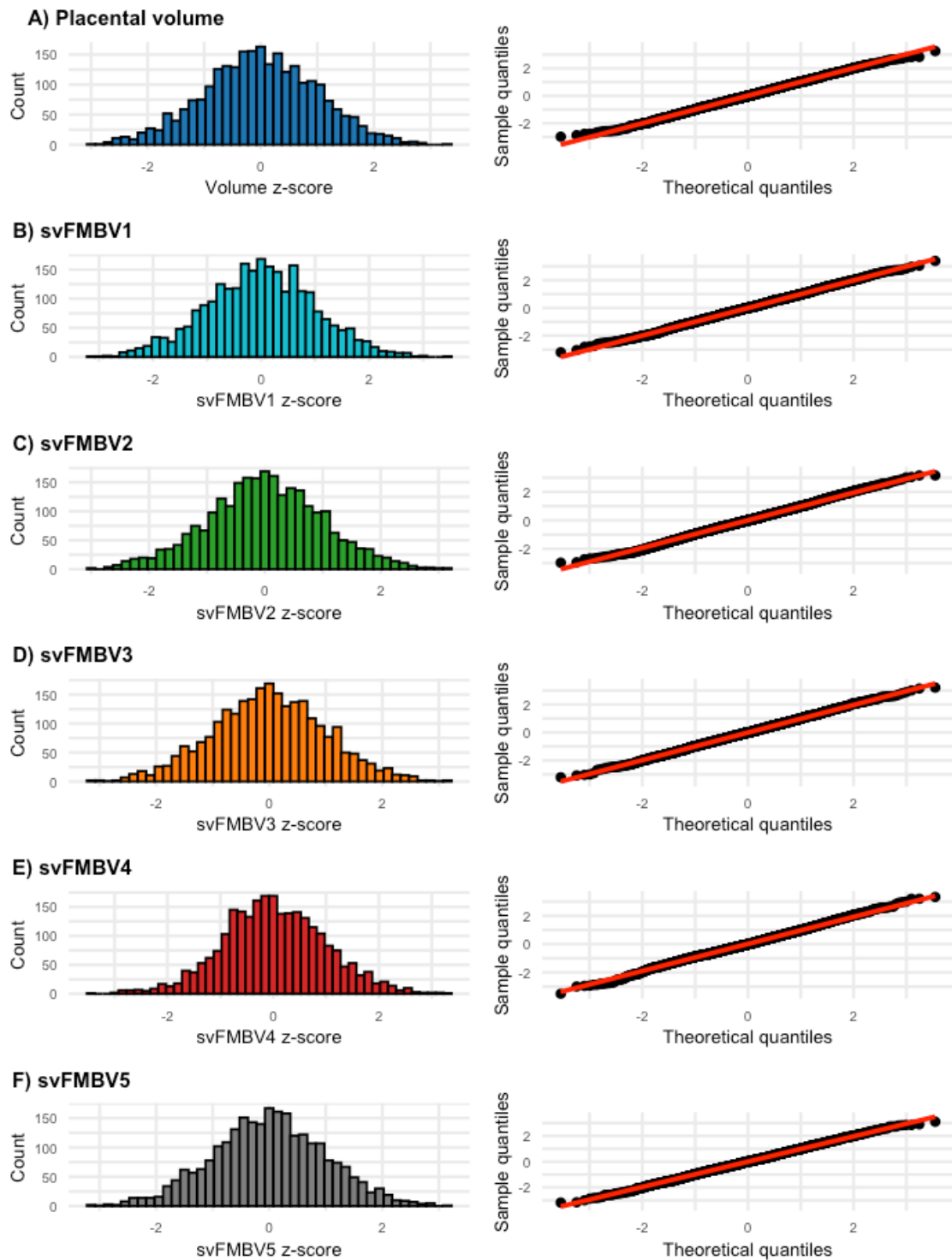
Figure 10 (page 79) displays histograms and Q–Q plots of the Z-scores for 3D-svFMBV1–5, calculated using their respective models. In every instance, the data demonstrated the expected normal distribution. These distributions were also derived in the development cohort and likewise describe apparent fit.

Table 11: Percentage of cases at various centiles observed in study cohort (Obs) vs predicted by final model (Pred) for A) placental volume, svFMBV1 and svFMBV2; and B) svFMBV3, svFMBV4, and svFMBV5

A)	Target centile	Volume		svFMBV1		svFMBV2	
		Pred (%)	Obs (%)	Pred (%)	Obs (%)	Pred (%)	Obs (%)
	1 st	1.18	1.02	0.94	1.02	1.02	1.02
	3 rd	3.38	3.02	3.61	3.02	3.26	3.02
	10 th	10.09	10.01	9.82	10.01	9.89	10.01
	25 th	24.85	25.01	25.13	25.01	24.30	25.01
	50 th	50.41	50.02	49.98	49.98	49.74	50.02
	75 th	74.79	74.99	75.38	74.99	75.58	74.99
	90 th	89.95	89.99	89.87	89.99	90.30	89.99
	97 th	97.02	96.98	96.90	96.98	96.70	96.98
	99 th	99.06	98.98	98.94	98.98	98.94	98.98

B)	Target centile	svFMBV3		svFMBV4		svFMBV5	
		Pred (%)	Obs (%)	Pred (%)	Obs (%)	Pred (%)	Obs (%)
	1 st	1.06	1.02	1.06	1.02	1.18	1.02
	3 rd	3.14	3.02	2.55	3.02	2.83	3.02
	10 th	10.33	10.01	8.32	10.01	10.48	10.01
	25 th	25.05	25.01	23.48	25.05	24.46	25.01
	50 th	50.65	50.02	50.45	50.02	49.55	50.02
	75 th	75.54	74.99	75.42	74.99	75.19	74.99
	90 th	90.50	89.99	90.22	89.99	89.75	89.99
	97 th	96.82	96.98	96.90	96.98	96.70	96.98
	99 th	98.94	98.98	98.98	98.98	98.98	98.98

Figure 10: Paired histograms and quantile-quantile plots for z-scores of each OxNNet-derived parameter

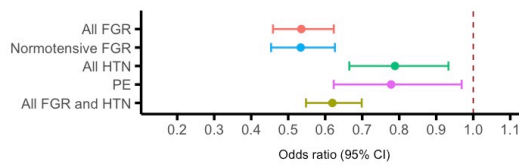


3.3.5 Prediction of adverse outcomes

Figure 11 (page 81) presents odds ratios and C-statistics for the prediction of adverse outcomes using the z-scores derived from each model for each OxNNet parameter. With the exception of the prediction of normotensive FGR from svFMBV2 and svFMBV4 z-score, all odds ratios demonstrated statistical significance (the 95% confidence intervals did not cross 1), with higher values being associated with lower risk of adverse outcome.

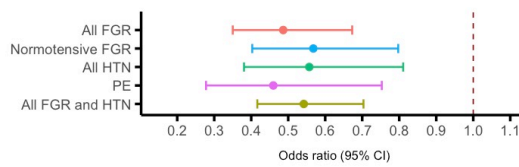
Figure 11: Forest plots, odds ratios and C-statistics for the prediction of adverse perinatal outcomes in the complete FirstPLUS cohort using z-scores for each OxNNet parameter

Prediction of adverse outcomes with placental volume z-score



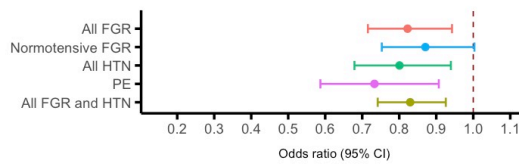
OR (95% CI)	C-statistic (95% CI)
0.54 (0.46–0.62)	0.65 (0.62–0.69)
0.53 (0.45–0.63)	0.65 (0.61–0.69)
0.79 (0.67–0.93)	0.57 (0.53–0.61)
0.78 (0.62–0.97)	0.58 (0.52–0.63)
0.62 (0.55–0.70)	0.62 (0.59–0.65)

Prediction of adverse outcomes with svFMBV1 z-score



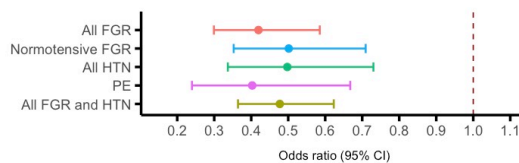
OR (95% CI)	C-statistic (95% CI)
0.49 (0.35–0.67)	0.58 (0.54–0.62)
0.57 (0.40–0.80)	0.56 (0.52–0.60)
0.56 (0.38–0.81)	0.56 (0.52–0.61)
0.46 (0.28–0.75)	0.59 (0.53–0.65)
0.54 (0.42–0.70)	0.57 (0.54–0.60)

Prediction of adverse outcomes with svFMBV2 z-score



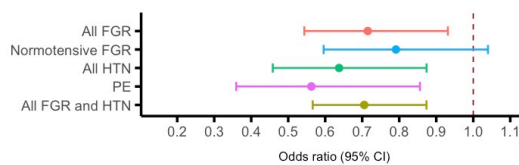
OR (95% CI)	C-statistic (95% CI)
0.82 (0.72–0.94)	0.56 (0.52–0.60)
0.87 (0.75–1.00)	0.54 (0.50–0.58)
0.80 (0.68–0.94)	0.56 (0.52–0.61)
0.73 (0.59–0.91)	0.59 (0.53–0.65)
0.83 (0.74–0.93)	0.55 (0.52–0.59)

Prediction of adverse outcomes with svFMBV3 z-score



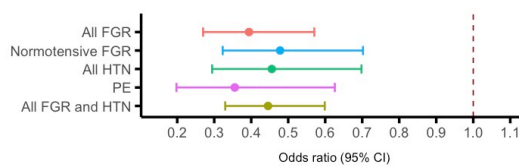
OR (95% CI)	C-statistic (95% CI)
0.42 (0.30–0.59)	0.60 (0.56–0.63)
0.50 (0.35–0.71)	0.57 (0.53–0.62)
0.50 (0.34–0.73)	0.57 (0.53–0.62)
0.40 (0.24–0.67)	0.60 (0.54–0.66)
0.48 (0.36–0.62)	0.58 (0.55–0.61)

Prediction of adverse outcomes with svFMBV4 z-score



OR (95% CI)	C-statistic (95% CI)
0.72 (0.54–0.93)	0.56 (0.52–0.59)
0.79 (0.60–1.04)	0.54 (0.50–0.58)
0.64 (0.46–0.87)	0.56 (0.52–0.61)
0.56 (0.36–0.86)	0.59 (0.53–0.65)
0.71 (0.57–0.87)	0.56 (0.53–0.59)

Prediction of adverse outcomes with svFMBV5 z-score



OR (95% CI)	C-statistic (95% CI)
0.39 (0.27–0.57)	0.59 (0.56–0.63)
0.48 (0.32–0.70)	0.57 (0.53–0.61)
0.46 (0.29–0.70)	0.58 (0.53–0.62)
0.36 (0.20–0.63)	0.60 (0.55–0.66)
0.45 (0.33–0.60)	0.58 (0.55–0.61)

3.4 Discussion and conclusion

With this work, novel reference ranges have been established for first trimester placental volume, and svFMBV at five locations within the uteroplacental vasculature. All models have demonstrated good apparent calibration in the cohort in which they were developed, with the exception of svFMBV4 in the lower tail, although they have not yet been externally validated, and it has further been shown that the resulting z-scores have potential utility in the prediction of adverse perinatal outcomes. The availability of robust reference charts for these OxNNet-derived parameters provides a foundation for future studies investigating placental development and its relationship to pregnancy outcome.

3.4.1 Comparison with previous work

There have been previous attempts to develop reference charts for first trimester placental volume, but they have significant methodological limitations. With a cohort of 236 low-risk pregnancies and using VOCAL™ (GE Healthcare) for placental segmentation and volume calculation, Meengeonthong *et al.* (2016) used the mean and standard deviation approach to create a reference chart for first trimester placental volume.¹⁵² Although the included cases are described as low-risk, very little demographic information is provided making it challenging to assess generalisability, and such a small cohort is unlikely to have statistical power to achieve modelling precision at the upper and lower extremes. A feasibility study by Guyomard *et al.* (2013) also used VOCAL™ for placental segmentation and volume calculation in 128 cases.¹⁵⁴ Again, concerns about the limitations of this small sample size apply. Once again, mean and standard deviation-based modelling was used, but in this case the standard deviation was treated as a constant. This approach is unlikely to have the flexibility to achieve modelling precision across the entire gestational age range, especially given the observation that distribution of placental volume values changes with gestational age — a finding which is supported by the studies discussed here.

The reliance on VOCAL™ in both of these studies is also a significant limitation because of concerns about inaccuracy of volume calculation with respect to irregular shapes, including the placenta which

is often not uniform in the first trimester,^{106,107} and comparatively poor performance with respect to reproducibility.¹⁰⁸

The largest study to date is that of Trilla Solà *et al.* (2024).¹⁵³ In this study of 1142 uncomplicated pregnancies, placental volume was derived from manual segmentation, generally considered to be the gold standard. Regression modelling of the mean and standard deviation was used to establish reference values by completed weeks of gestation, and the threshold volumes corresponding to the 5th, 10th, 50th, 90th and 95th centiles are provided. However, the coefficients needed for calculating z-scores or centiles from raw measurements have not been made available, which restricts clinical and research utility.¹⁵³ Further, as discussed above (section **3.2.2**), even this larger sample size may not be sufficient to achieve modelling precision at extreme centiles, and the authors do not comment on any power calculations or validation of their assumptions in this respect.

At the time of writing, no other attempts to establish centile charts for uteroplacental FMBV were identified. However, using the semi-automated random walker method of placental segmentation in 143 cases (including cases of PE and SGA as well as healthy cases), Collins *et al.* (2017) reported FMBV values ranging from 0.01 to 0.65 at the UPI (equivalent to 3D-svFMBV1 in this study) and <0.01 to 0.44 in the intervillous space (equivalent to 3D-svFMBV2 in this study).¹²⁵ This is a much narrower range of observations compared with this dataset, but these differences could be explained by substantial disparity in sample size.

3.4.2 Strengths and limitations

One of the principal strengths of this work lies in the large sample size. With more than twice the number of cases compared to the largest prior study, the dataset captures a broader spectrum of physiological variation, enhances representation of subgroups, and reduces the influence of outliers. This scale substantially improves the reliability of centile calculations, particularly at the extremes (e.g., the 5th and 95th percentiles). Rigorous quality control of imaging and segmentation further ensured that the models were derived from accurate placental assessments.

A further strength is the adoption of flexible modelling approaches capable of accommodating complex data distributions. In particular, the use of the Lambda-Mu-Sigma method enabled

construction of centile models that account for non-normal distributions and heteroscedasticity across gestation. This represents a significant advance over traditional mean and standard deviation approaches, yielding more accurate and clinically meaningful centile curves. Model complexity was constrained by comparing candidate distributions using the Akaike Information Criterion, which penalises the addition of parameters, alongside visual inspection of the fitted curves. This guards against unnecessary complexity, but it does not by itself establish that the models generalise beyond the sample in which they were developed.

Reproducibility was also enhanced through automated segmentation with OxNNet. This approach overcomes the limitations of manual or geometric methods, which may misrepresent the irregular morphology of the placenta, especially in early pregnancy.

Nevertheless, several limitations must be acknowledged. Representation at the extremes of gestational age was limited, reducing confidence in distributional estimates at very early and very late time points. The study was conducted at a single UK centre, which may constrain generalisability to populations with different demographic profiles or imaging systems. Although cases with overt pathology were excluded, undiagnosed or subclinical conditions may still have influenced measurements. The Lambda-Mu-Sigma method, while offering coherent conditional distributions and non-crossing centiles, introduces additional complexity to percentile calculation, such that statistical expertise or bespoke digital tools are required for implementation. To mitigate this barrier, models have been made available as R project files alongside publications arising from this work, facilitating replication and wider use. Finally, although OxNNet is designed to be vendor-agnostic through NIfTI conversion, multicentre validation across diverse ultrasound platforms remains necessary to confirm generalisability. Most importantly, the centile charts presented in this chapter have been assessed only in the cohort in which they were developed. The centile calibration and z-score distributions reported above are measures of apparent fit, and are optimistic by construction because the same observations were used both to fit and to assess the models. No external validation of these charts has been undertaken. Application to an independent dataset, most immediately the OxPLUS cohort and subsequently to populations scanned on other platforms, is required before the centiles can be relied upon for clinical interpretation.

3.4.3 Clinical and Research Implications

The utility of these reference charts lies not in binary classification of placental function as normal or abnormal, but in their integration into multivariate predictive models for adverse pregnancy outcomes. Reliable centile charts for both perfusion and volume provide a quantitative framework for risk stratification, supporting early identification of pregnancies at risk of complications such as pre-eclampsia or fetal growth restriction. By enabling reproducible, automated measurement, OxNNet makes large-scale application feasible in both research and clinical practice.

3.4.4 Conclusion

In summary, this work provides comprehensive reference ranges for first-trimester placental perfusion and volume, derived from a large, rigorously curated cohort using advanced machine learning and statistical modelling. Automated segmentation and LMS-based centile construction provide robust, reproducible charts that improve upon previous work. While limitations remain, particularly regarding generalisability and representation at extreme gestational ages, these findings establish a valuable foundation for future research and clinical application. Together, they advance the quantitative assessment of placental function in early pregnancy and support the development of predictive models for adverse outcomes.

[4] DEVELOPMENT OF TRAINING AND QUALITY CONTROL PROCEDURES

4.1 Introduction

The necessity of thorough training in any ultrasound procedure is evident, especially when it comes to pregnancy sonography, and the importance of this subject is emphasised in guidance and policies of numerous professional bodies including the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG), the Society for Maternal-Fetal Medicine (SMFM), and the World Health Organization (WHO).¹⁶⁴⁻¹⁶⁷

To commence the OxPLUS study, it was necessary to train all participating sonographers in 3D first trimester placental ultrasonography — a technique that was previously unfamiliar to all practitioners. In addition to fulfilling the study's primary aims, this process enabled the collection of data regarding the extent of training necessary to integrate this innovative method into a standard NHS maternity ultrasound setting. Although an experienced sonographer will have skills that are directly transferable to placental sonography, there are also several potential pitfalls which could render an image unusable or could result in the generation of inaccurate measurements and unreliable data. Therefore, in addition to training, ongoing quality control (QC) is extremely important.

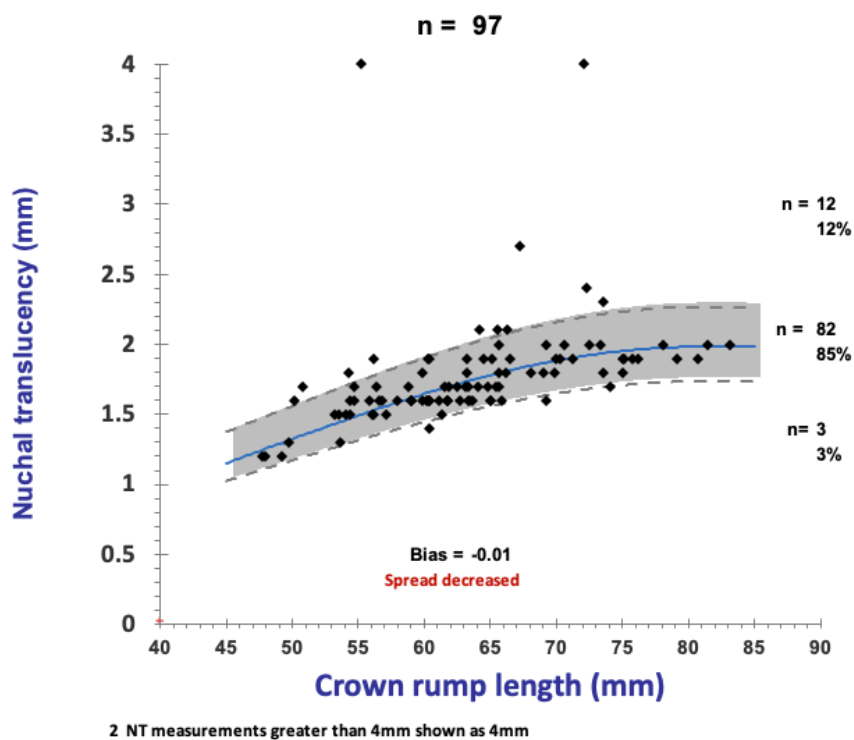
Sonographers' images and measurements are routinely audited for QC purposes. For example, first trimester nuchal translucency (NT) measurement in the NHS Fetal Anomaly Screening Programme involves regular self-assessment using a feedback tool (Figure 12, page 88).^{168,169} This tool uses a 'traffic light' warning system to indicate measurement bias and spread, highlighting outliers. While bias does not always mean poor performance, it allows sonographers to reflect on potential learning needs and to pursue further training if needed. Familiarity with this feedback approach, and its similarity to first trimester biomarker use for risk stratification, makes it a fitting QC model for first trimester 3D placental ultrasound with power Doppler.

The challenge with a novel technique or marker lies in defining quality control failures. Some QC failures may be obvious — such as failure to obtain a measurement — but others may be more nuanced, such as those based on identification of statistical outliers or data trends. Criteria that are

too stringent may result in false positive QC 'alarms', whereas if they are too lenient then a greater quantity of data will be required to trigger the QC warning, potentially allowing problematic practice to continue for longer. Some degree of trial and error is inevitable as QC systems are calibrated, particularly when there is little known about the expected performance of a new test or technique, but there are statistical approaches which may offer insight including the Westgard rules and cumulative sum (CUSUM) analysis.

Figure 12: Example of a diagnostic plot self-assessment tool for nuchal translucency measurements.

Image from NHS fetal anomaly screening programme website, produced by Prof. Dave Wright, University of Exeter.¹⁶⁹



4.1.1 Westgard rules

The Westgard rules, also known as ‘multi-rule quality control’, are a combination of decision criteria or ‘control rules’ which form a set of statistical patterns unlikely to occur by chance (**Table 12**). Developed for use in laboratory quality control, these rules help to determine whether an analytical run is in-control or out-of-control. In this context, in-control means that the measures are within the expected degree of random variation. Out-of-control means that one or more rules have been breached, suggesting a non-random problem which — in a laboratory context — might include reagent degradation, poor instrument calibration, or operator error etc. These rules are designed to maximise error detection (true positives) while keeping erroneous rejections to a minimum (false positives). An adaptation of these rules may help to provide a framework for evaluation of sonographer performance with respect to first trimester 3D placental ultrasound with power Doppler.

Table 12: Westgard multi-control rules

Rule name	Description	Interpretation
1 2s	One control result exceeds ± 2 standard deviations (SD) from the mean	Warning only
1 3s	One control result exceeds ± 3 SD	Random or systematic error
2 2s	Two consecutive results exceed ± 2 SD on the same side of the mean	Systematic error
R 4s	One result exceeds $+2$ SD and another -2 SD in the same run	Random error
4 1s	Four consecutive results exceed ± 1 SD on the same side	Systematic error
10 x	Ten consecutive results fall on the same side of the mean	Systematic error

4.1.2 Cumulative Sum (CUSUM) methodology

The Cumulative Sum (CUSUM) method complements multi-rule approaches by tracking performance over time and quantifying learning curves and shifts in operator performance. CUSUM is particularly well suited to sonography because it accumulates small deviations that, when persistent, indicate a meaningful change in performance; this makes it sensitive to gradual deterioration or improvement and useful for determining when retraining or supervision is required. Published applications in fetal biometry, nuchal translucency and other sonographic tasks demonstrate that CUSUM can identify the number of cases required to reach competence, highlight periods of systematic bias, and support targeted remediation while minimising unnecessary interventions for random variation.^{170,171}

4.1.3 Summary

In summary, the development of training and quality control processes for 3D placental ultrasound with power Doppler requires a multifaceted approach: structured training, objective assessment tools, clear definitions of quality control failure, and ongoing performance monitoring.

4.1.4 Aims

- To design and implement a structured self-assessment feedback tool for sonographers performing first trimester three-dimensional placental ultrasound with power Doppler.
- To describe changes in sonographer performance before and after feedback is received.
- To establish objective criteria for quality control failure in placental ultrasound by adapting the Westgard statistical rules, and to longitudinally monitor individual sonographer performance using cumulative sum (CUSUM) analysis, thereby identifying learning curves and periods of suboptimal practice.

4.1.5 Statement of personal contribution

Direct supervision of sonographers in training was undertaken by me, in conjunction with Theophilus Adu-Bredu (research sonographer). All subsequent data collection and analysis was performed independently.

4.2 Methodology

4.2.1 Sonographer training protocols for OxPLUS

Prior to commencement of the OxPLUS study, a training package was developed. Sonographers were provided with an instructional video describing the process of placental capture with 3D ultrasound and power Doppler, as described in section **2.2.7.4**, along with a detailed standard operating procedure document which they were required to read. Quick-reference summary posters were also placed in each scan room at initiation of recruitment.

Sonographers were initially required to perform scans under supervision of a practitioner experienced in first trimester 3D placental ultrasound with power Doppler. In each case, an individual sonographer training record was maintained with a record of guidance or intervention needed for each scan, and the signature of the trainer was required once each scan was complete. Once three consecutive scans had been performed without a need for intervention or guidance, the training record was reviewed by the Principal Investigator, and the sonographer was 'signed off' as competent to perform the scans independently. After this point, supervision was no longer mandated but remained available to sonographers upon request for the duration of the study.

The threshold of three scans without intervention was decided pragmatically, based on an assumption that this level of supervision would likely be deliverable both in the context of this study and in any future clinical implementation; and was also modelled on the requirements of the Fetal Medicine Foundation certification in pre-eclampsia screening, which includes competence in first trimester uterine artery Doppler velocimetry and requires submission of three high quality images which pass quality control in order to confirm competence.¹⁷²

4.2.2 Development and delivery of feedback tool

To construct OxNNet parameter feedback tools, early iterations of the centile charts described in chapter 3 were used. Two tools were developed, one for placental volume and another for svFMBV2 — the latter was selected pragmatically over other svFMBV measures because at the time the comparative predictive value of each measure was not known, and FMBV at the intervillous space (equivalent to svFMBV2) has been shown to have an association with pre-eclampsia.¹²⁵ The method used to construct the preliminary charts is identical to the description of the LMS methodology in section 3.2.4, except that the entire FirstPLUS cohort was used (the derivation of which is shown in **Figure 3**, page 52), and only the default GAMLSS families were explored (including Box–Cox Cole and Green original, Box–Cox Power Exponential original, and Box–Cox t original distributions). The Box–Cox Power Exponential original distribution was selected for modelling of both parameters. A further difference is that placental volume was presented in millimetres cubed, rather than centimetres cubed.

A script was prepared in R to ensure consistency of feedback across all sonographers. Each chart included a centile line for the 50th centile, and fans for the 25th–75th, and the 5th–95th centiles. Individual sonographer measurements were extracted from the working OxPLUS dataset and were plotted on the chart. Numerical values for bias, spread, and failed placental segmentation were also presented.

4.2.2.1 Calculation of bias

Bias was calculated as the mean of the residuals for the individual sonographer, and these values were compared to the expected values (median) derived from the reference dataset (FirstPLUS). Advisory and corrective comments were triggered if the absolute value of the bias exceeded 0.1 or 0.2 times the reference median respectively; the absolute value was used so that systematic under-estimation and over-estimation are treated alike. A bias of 0.1 times the median corresponds to a multiple of the median of between 0.9 and 1.1 (**Table 13**, page 94). Additionally, the percentage of results falling within the expected top 25%, middle 50%, and bottom 25% was presented, providing an intuitive and familiar way for sonographers to understand the direction of any bias identified.

4.2.2.2 Calculation of spread

Spread was calculated as the standard deviation of the residuals for an individual sonographer, and these were compared with the mean standard deviation of measurements in the FirstPLUS dataset. Advisory and corrective comments were triggered if the calculated spread exceeded the reference value by more than 10% or 20% respectively, that is, if it exceeded 1.1 or 1.2 times the reference standard deviation (Table 13, page 94).

4.2.2.3 Count of segmentation failures

The feedback report also included a count of scans where placental segmentation or svFMBV2 processing had failed, which is most likely to reflect poor image quality or missing data (i.e. a placental scan was performed but no power Doppler signal was identified). A placental volume result was counted as a segmentation failure if it was missing or below 10 cm³, and an svFMBV2 result was counted as a failure if it was missing or below 0.01; the two were counted separately. An advisory comment was triggered by a single failure, and a corrective comment by more than one.

Table 13: Conditional logic and resulting textual outputs for self-assessment feedback tool for placental volume and svFMBV2

Parameter	Logic	Textual output
Bias value	IF (abs(bias) <= 0.1 * reference_median)	This result is within 0.1 MoMs, no action is required.
	ELSE IF (abs(bias) <= 0.2 * reference_median)	This is acceptable but could be improved (within 0.2 MoMs). Please review the SOP.
	ELSE IF (abs(bias) > 0.2 * reference_median)	This suggests need for improvement (>0.2 MoMs). Please arrange further training.
Volume bias sign	IF (vol_bias > 0)	Positive bias may be due to over segmentation of the placenta. Remember to adjust greyscale gain to distinguish placenta from myometrium.
	ELSE IF (vol_bias < 0)	Negative bias suggests the placenta is sometimes incomplete or in shadow. Please ensure you have captured the entire placenta.
	ELSE	NA
FMBV bias sign	IF (fmbv_bias > 0)	Positive bias may be due to high PD gain. Please ensure you select the subnoise gain setting correctly.
	ELSE IF (fmbv_bias < 0)	Negative bias may be due to low PD gain. Please ensure you select the subnoise gain setting correctly.
	ELSE	NA
Spread	IF (spread <= 1.1 * reference_SD)	This is <10% difference from the expected result. The spread of your results is as expected. No action is required.
	ELSE IF (spread <= 1.2 * reference_SD)	This is acceptable but could be improved (10-20% difference from expected result). There is slightly more variability in your results than expected. Please review the SOP.
	ELSE IF (spread > 1.2 * reference_SD)	This suggests need for improvement (>20% difference from expected result). There is more variability in your results than expected. Please arrange further training.
Segmentation failure	SUM(segmentation_failure) = 0	No cases of failed segmentation were identified.
	SUM(segmentation_failure) = 1	Placental segmentation failed in one of your images. Please review the SOP.
	SUM(segmentation_failure) > 1	Placental segmentation failed in some of your images. Please arrange further training.
FMBV failure	SUM(fmbv_failure) = 0	No cases of failed FMBV assessment were identified.
	SUM(fmbv_failure) = 1	FMBV assessment failed in one of your images. Please review the SOP.
	SUM(fmbv_failure) > 1	FMBV assessment failed in some of your images. Please arrange further training.

4.2.3 Delivery of feedback and corrective action

Recruitment for the OxPLUS study, and consequently training of sonographers, commenced in April 2023 and concluded in August 2024. Individual performance was audited using the OxNNet feedback toolkit after 1 year of recruitment, in April 2024.

4.2.3.1 Power calculation

A power calculation was performed using the `pwr.t.test` function from the `pwr` package in R to determine the number of scans required to detect bias. Assuming Cohen's *d* of 0.4 (a moderate effect size), a significance level of 0.05 and power of 0.7, then 41 cases would be required. The choice of 50 cases was pragmatic, to allow for some failures of segmentation.

4.2.3.2 Sonographer inclusion criteria

Sonographers were eligible for inclusion if they had completed training in first trimester 3D placental ultrasound with power Doppler and had performed at least 50 scans by April 2024. The most recent 50 scans were selected for each sonographer, and the measurements entered into the OxNNet feedback toolkit.

4.2.3.3 Feedback and training protocol

Feedback outputs were generated for both placental volume and svFMBV2, and these were provided to the sonographers. Where minor deviations were identified, a discussion took place with the principal investigator to ensure that the potential causes of bias were understood, and sonographers were asked to revisit the standard operating procedure document. Where need for improvement was identified, the initial training protocol was repeated as described in paragraph 4.2.1.

4.2.3.4 Comparison of performance before and after feedback

For those sonographers who received feedback, their scans for the remaining months of the study were evaluated in the same manner. Sonographers with fewer than 29 evaluable cases after feedback were excluded from further analysis. This threshold was selected based on the power calculation described in paragraph 4.2.3.1, modified to assume a Cohen's *d* statistic of 0.5 (a moderate effect size, but slightly greater than the initial assumption of 0.4) which determined that 27 cases should be sufficient, and this was inflated to 29 to account for expected image processing failures.

Sonographer performance before and after retraining was compared in those cases that remained eligible for inclusion. For each parameter (placental volume and svFMBV2) each metric (bias, spread and processing failure) was converted to a categorical outcome of pass, advisory, or fail (as described in section 4.2.2). Bhapkar's test was used for statistical comparison. svFMBV2 processing failures were expressed as a percentage of the total number of scans by that sonographer, and these proportions were compared before and after feedback using a paired Wilcoxon signed rank test. Failures of placental segmentation were not evaluated separately, since all segmentation failures were also svFMBV2 failures, but some svFMBV2 failures were not volume failures.

4.2.4 Multi-rule quality control

Upon the conclusion of the OxPLUS study, data were grouped according to the ultrasound practitioner performing the scan and were analysed retrospectively.

4.2.4.1 Quality control failure triggers

Each case was evaluated against 11 binary (pass or fail) criteria modelled on the Westgard multi-rule QC process (**Table 14**, page 97). The z-scores for placental volume and svFMBV5 were required for QC, and these were calculated according to the published methods described in chapter 3.^{173,174} A decision was taken not to include an assessment of all five svFMBV parameters, but rather to focus on svFMBV5 measurements only on the basis that this appears to have the strongest association with adverse outcomes (as discussed in paragraph 3.3.5).

Using these criteria, we defined three failure categories. Definition 1 encompassed all failure parameters. Definition 2 included only cases with a correctly saved image, and definition 3 included only cases in which an image could be successfully processed. We examined Definitions 2 and 3 because failure to save an image does not necessarily indicate lack of technical competence; excluding unsaved images therefore gives a clearer measure of the operator's sonographic skill. Definition 3 is motivated by OxNNet's intended real-time use during scanning: the system will prompt the operator to repeat acquisition if no image is received in the correct format. Excluding these events from the failure definition better reflects expected real-world performance. However, failure of OxNNet's internal QC process (which checks that the placenta has not been incompletely captured at the edge of the image) was not a reason for exclusion because this does reflect sonographer competence.

4.2.4.2 Cohort definitions

To analyse each failure definition, three distinct cohorts were examined. Cohort 1 comprised all cases within OxPLUS in which a placental scan was performed. Cohort 2 excluded instances where scan images were either not saved or were stored in an incorrect format. Cohort 3 consisted solely of cases in which OxNNNet parameters were processed successfully.

Table 14: Multi-rule approach to quality control

Reason for failure	Failure definitions		
	1	2	3
Images not saved or saved in wrong format	Included	Excluded	Excluded
Processing failure	Included	Included	Excluded
OxNNNet internal QC failure	Included	Included	Included
Placental volume z-score exceeds ± 3	Included	Included	Included
svFMBV5 z-score exceeds ± 3	Included	Included	Included
Placental volume z-score exceeds ± 2 on the same side of the mean for second consecutive case	Included	Included	Included
svFMBV5 z-score exceeds ± 2 on the same side of the mean for second consecutive case	Included	Included	Included
Placental volume z-score exceeds ± 1 on the same side of the mean for fourth consecutive case	Included	Included	Included
svFMBV5 z-score exceeds ± 1 on the same side of the mean for fourth consecutive case	Included	Included	Included
Placental volume z-score on the same side of the mean for tenth consecutive case	Included	Included	Included
svFMBV5 z-score on the same side of the mean for tenth consecutive case	Included	Included	Included

4.2.5 Cumulative Sum (CUSUM) learning-curve analysis

A learning-curve Bernoulli CUSUM was constructed for each sonographer to monitor sequential performance across all OxPLUS scans performed by that individual.¹⁷⁵⁻¹⁷⁷ The event of interest was successful image acquisition with no QC triggers, and the values for ξ_i (the binary indicator that is positive when the i -th case shows the event of interest) were coded as one and zero for pass and fail, respectively.

The baseline success probability p_0 was estimated as the sample mean of ξ , and this was calculated separately for each cohort (**Table 15**). These values are the complement of the composite quality-control failure rates for each cohort.

Table 15: Values of p_0 and h used for CUSUM analysis

Cohort	p_0	h
1	0.72	3.63
2	0.74	3.33
3	0.79	2.76

To determine reassurance, warning, and alarm thresholds, alternative success probabilities p_1 were specified using relative multipliers: an alert shift of $p_1 = p_0 \times 0.90$ (10% relative shift) was used.

Per-case log-likelihood increments were computed as shown in **Equation 2** and the CUSUM calculated as a plain cumulative sum (without truncation, where positive and negative excursions were interpreted according to the coding of ξ) according to **Equation 3**, where $\text{CUSUM}i$ is the cumulative sum of the case scores up to the i -th case. That is, $\text{CUSUM}i = s_1 + s_2 + \dots + s_i$, with each s_j the log-likelihood increment contributed by case j (computed from the observed outcome ξ_j and the baseline and alternative probabilities p_0 and p_1). Positive increments indicate improvement in performance; negative increments indicate deterioration.

Equation 2: Bernoulli log-likelihood increment

$$s_i = \log\left(\frac{p_0}{p_1}\right) \xi_i + \log\left(\frac{1-p_0}{1-p_1}\right) (1 - \xi_i)$$

Equation 3: Cumulative CUSUM statistic

$$\text{CUSUM}_i = \sum_{j \leq i} s_j.$$

The decision limit h was obtained by simulation using R. For each cohort, 10,000 in-control Bernoulli sequences of length $T = 100$ were simulated. The simulation was parameterised in terms of the failure rate, taken as the complement of p_0 , with an alternative failure rate 20% greater in relative terms rather than the success-based shift p_1 ; the highest CUSUM value reached in each simulated run was recorded, and h was set to the empirical $(1-\alpha)$ quantile of those peak values with $\alpha = 0.01$ (where α is the probability of false alarm). In plain terms, h is the value that only about 1% of in-control runs exceeded over the 100-case simulation horizon. The derived h values are presented in **Table 15** (page 98). Because the decision limit was calibrated on this failure-monitoring parameterisation and then applied to the learning-curve CUSUM, whose per-case increments are larger, the nominal 1% false-alarm rate should be regarded as approximate.

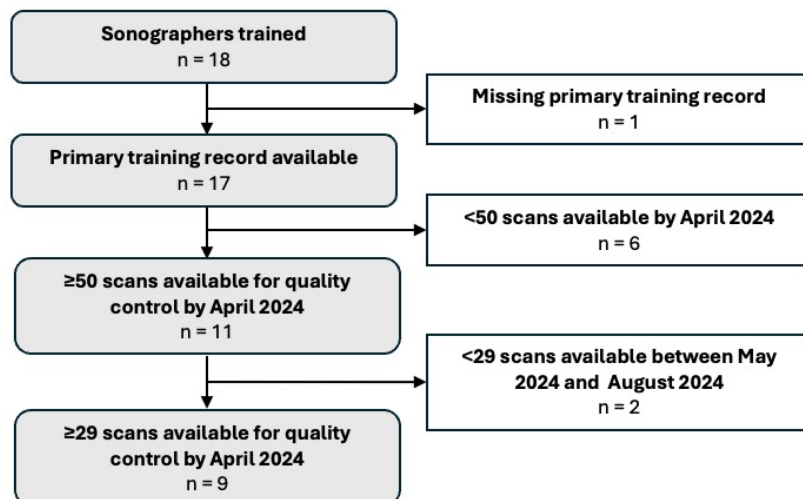
Charts were plotted per sonographer with superimposed reassurance, warning, and alarm thresholds. The reassurance threshold was set at $+h$, the warning threshold at $-h$, and the alarm threshold at $-2h$. Because a pass contributes a positive increment and a failure a negative one, a CUSUM rising to cross $+h$ indicates performance sustained above the baseline success rate, whereas a falling CUSUM crossing $-h$ and then $-2h$ indicates progressively greater cause for concern. A single value of $h = 3.33$ was applied to all three cohort analyses, the differences between cohorts being negligible, so that the thresholds shown in Figure 17 and in Appendices 1 and 2 are $+3.33$, -3.33 and -6.66 respectively. Additional vertical lines were added to indicate the date on which feedback was delivered using the feedback tool described in paragraph 4.2.2 (page 92). Where any corrective action was recommended, this line was colour-coded red. If the recommendation was only advisory, then the colour was amber, and if no corrective action or advisory statement was made then the line was green.

4.3 Results

In total, 18 sonographers were trained and deemed to achieve competence to perform first trimester 3D placental ultrasound with power Doppler during the OxPLUS study.

One sonographer is excluded from reports on the primary training process, since regrettably the training log was unavailable, although this did not necessitate exclusion from subsequent analyses. Of 18 potentially eligible sonographers, 11 had performed at least 50 scans by April 2024 and were eligible to take part in feedback using the novel toolkit. All 18 were included in the CUSUM learning curve analysis. A summary of these exclusions is shown in **Figure 13**.

Figure 13: Summary of sonographers included in training and quality control analyses



4.3.1 Primary training of sonographers

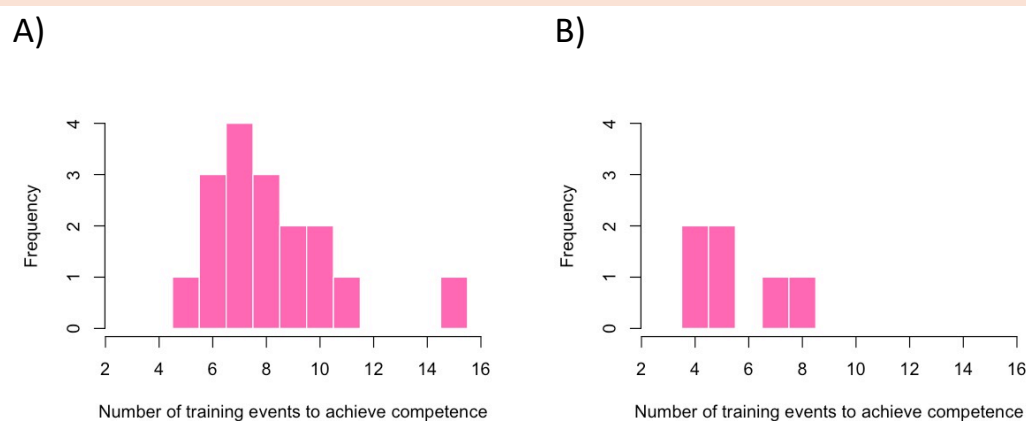
The median number of training events required to achieve competence, as judged at the time by the OxPLUS study team, was eight scans (with varying degrees of supervision), and a range of five to fifteen. A detailed breakdown of training events is presented in **Table 16**, and the distribution of training events is presented in **Figure 14**.

Table 16: Summary of learning events required to achieve competence in first trimester 3D placental ultrasound with power Doppler.

Primary training relates to training at commencement of the OxPLUS study, and secondary training relates to additional training delivered in response to the recommendations of the OxNNet feedback toolkit.

Type of training event	Median number of training events n (range)	
	Primary training n = 17	Secondary training n = 6
Demonstration	1 (1–4)	0
Guidance (hands on)	1 (0–2)	0
Guidance (hands off)	2 (0–10)	2 (1–5)
Unguided supervision	3	3
Total	8 (5–15)	5 (4–8)

Figure 14: Training events required to achieve competence in A) primary training in first trimester 3D placental ultrasound with power Doppler, and B) further training required in response to the OxNNet feedback toolkit

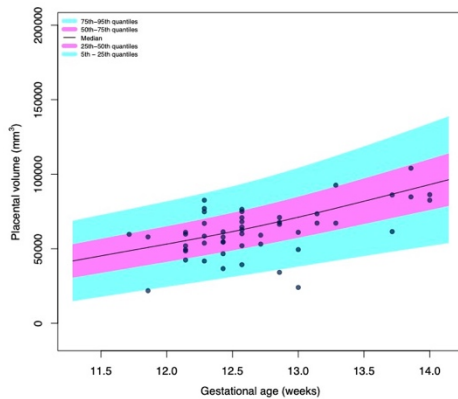


4.3.2 Implementation of the OxNNet feedback toolkit

Examples of the OxNNet feedback toolkit output, with cases that both passed and failed the QC process, are shown in **Figure 15**.

Figure 15: Examples of the self-assessment feedback tool for placental volume (A, B) and svFMBV2 (C, D) showing both positive and negative feedback

A) Placental volume, positive feedback

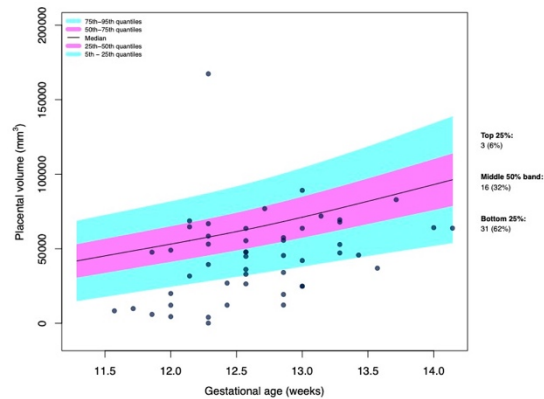


Bias: -3845 mm³
Bias is the mean difference between your measurements and the expected.
This result is within 0.1 Moku, no action is required.
Negative bias suggests the placenta is sometimes incomplete or in shadow. Please ensure you have captured the entire placenta.

Spread: 14065 mm³
Spread is the standard deviation of the difference between your measurements and the expected.
This is <15% difference from the expected result.
The spread of your results is as expected. No action is required.

Failed placental segmentation: 0 (0%)
No cases of failed segmentation were identified.

B) Placental volume, negative feedback

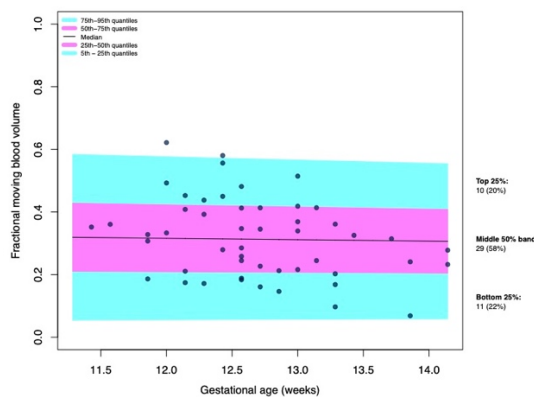


Bias: -20023 mm³
Bias is the mean difference between your measurements and the expected.
This suggests need for improvement (>0.2 Moku). Please arrange further training.
Negative bias suggests the placenta is sometimes incomplete or in shadow. Please ensure you have captured the entire placenta.

Spread: 27387 mm³
Spread is the standard deviation of the difference between your measurements and the expected.
This suggests need for improvement (>20% difference from expected result).
There is more variability in your results than expected. Please arrange further training.

Failed placental segmentation: 6 (12%)
Placental segmentation failed in some of your images. Please arrange further training.

C) svFMBV2, positive feedback

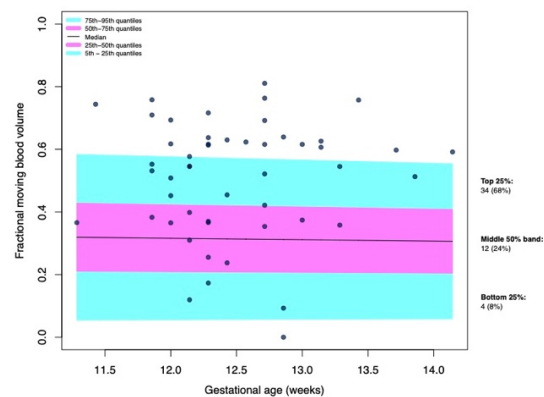


Bias: 0.0032
Bias is the mean difference between your measurements and the expected.
This result is within 0.1 Moku, no action is required.
Positive bias may be due to high PD gain. Please ensure you select the subnoise gain setting correctly.

Spread: 0.13
Spread is the standard deviation of the difference between your measurements and the expected.
This is <15% difference from the expected result.
The spread of your results is as expected. No action is required.

Failed FMBV assessment: 0 (0%)
No cases of failed FMBV assessment were identified.

D) svFMBV2, negative feedback



Bias: 0.19
Bias is the mean difference between your measurements and the expected.
This suggests need for improvement (>0.2 Moku). Please arrange further training.
Positive bias may be due to high PD gain. Please ensure you select the subnoise gain setting correctly.

Spread: 0.19
Spread is the standard deviation of the difference between your measurements and the expected.
This suggests need for improvement (>20% difference from expected result).
There is more variability in your results than expected. Please arrange further training.

Failed FMBV assessment: 1 (2%)
FMBV assessment failed in one of your images. Please review the SOP.

Feedback was delivered to all 11 eligible sonographers, of whom nine had sufficient scans in the post-feedback period to permit paired comparison. For these nine, the feedback given can be categorised into three categories — pass, fail, or advisory — for each metric evaluated (**Table 17**, page 104). The commonest reason for QC failure was detection of bias in svFMBV2 measurement, which was identified in 66.7% (n = 6) of the cases evaluated. Performance was better with respect to placental volume, with only one case failing QC on this basis, although an advisory comment suggesting some need for improvement was delivered in 66.7% of cases (n = 6). When the toolkit was applied retrospectively to all scans performed by each sonographer for the remainder of the study, a significant difference was noted in placental volume bias classification ($P < 0.001$), with 44% (n = 4) showing improvement. However, the performance of two sonographers worsened after feedback, with 33% (n = 3) being in the failure category. No other metric demonstrated a statistically significant difference, but the low case numbers may limit the ability to detect change and could magnify differences that have occurred by chance.

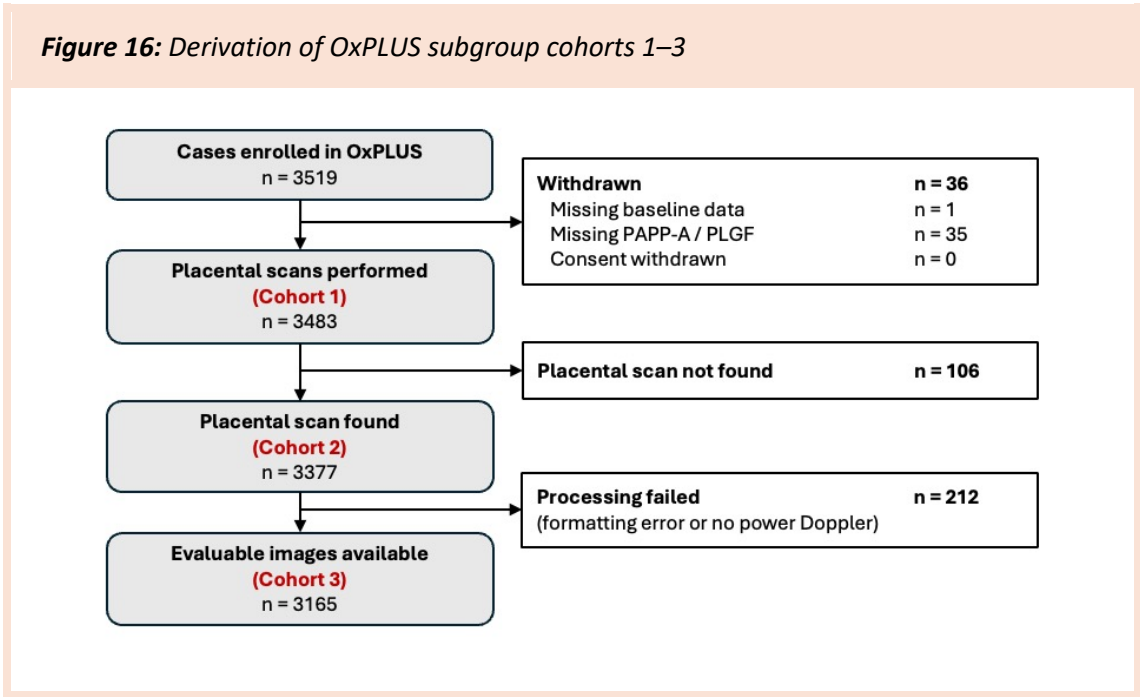
Table 17: Sonographer performance as rated using the OxNNet feedback toolkit before and after receiving feedback

Metric	Before feedback (n = 9) n (%)	After feedback (n = 9) n (%)	P
Volume bias classification			
Pass	2 (22.2%)	5 (55.6%)	<0.001
Advisory	6 (66.7%)	1 (11.1%)	
Fail	1 (11.1%)	3 (33.3%)	
Change in volume bias classification			
Improved		4 (44.4%)	
Unchanged		3 (33.3%)	
Worsened		2 (22.2%)	
Volume spread classification			
Pass	7 (77.8%)	7 (77.8%)	1
Advisory	0	0	
Fail	2 (22.2%)	2 (22.2%)	
Change in volume spread classification			
Improved		1 (11.1%)	
Unchanged		7 (77.8%)	
Worsened		1 (11.1%)	
FMBV bias classification			
Pass	3 (33.3%)	2 (22.2%)	0.556
Advisory	0	0	
Fail	6 (66.7%)	7 (77.8%)	
Change in FMBV bias classification			
Improved		1 (11.1%)	
Unchanged		6 (66.7%)	
Worsened		2 (22.2%)	
FMBV spread classification			
Pass	6 (66.7%)	7 (77.8%)	0.277
Advisory	0	1 (11.1%)	
Fail	3 (33.3%)	1 (11.1%)	
Change in FMBV spread classification			
Improved		2 (22.2%)	
Unchanged		7 (77.8%)	
Worsened		0	

4.3.3 Multi-rule quality control and CUSUM learning curve analysis

4.3.3.1 Derivation of the cohorts 1-3

A summary of the derivation of the three subgroup cohorts from the OxPLUS cohort is shown in **Figure 16**. After exclusions, there were 3483 cases in cohort 1, 3377 in cohort 2, and 3165 in cohort 3.



4.3.3.2 Exploration of multi-rule QC failures

The incidence of multi-rule QC failures across all three cohorts is presented in **Table 18** (page 106). The commonest reason for failure by some distance was the Z-score being on the same side of the mean for the tenth consecutive case, which reflects systematic bias in measurement, and this was true across all cohorts with incidence of 9.13%, 9.42%, and 10.05% in cohorts 1–3 respectively. The next most common reasons for failure were processing failures and failure of OxNNet’s internal QC process.

Table 18: Summary of the incidence of multi-rule quality control failures across subgroup cohorts 1–3

Failure triggers	n	Incidence of failures in each cohort (%)		
		Cohort 1 All cases n = 3483	Cohort 2 Scan found n = 3377	Cohort 3 Scan processed n = 3165
Image not saved or wrong format	106	3.04		
Processing failure	212	6.09	6.28	
OxNNet internal QC failure	148	4.25	4.38	4.68
Z-score exceeds ± 3				
Placental volume	90	2.58	2.67	2.84
svFMBV5	48	1.38	1.42	1.52
Z-score exceeds ± 2 on the same side of the mean for second consecutive case				
Placental volume	27	0.78	0.80	0.85
svFMBV5	68	1.95	2.01	2.15
Z-score exceeds ± 1 on the same side of the mean for fourth consecutive case				
Placental volume	19	0.55	0.56	0.60
svFMBV5	111	3.19	3.29	3.51
Z-score on the same side of the mean for tenth consecutive case				
Placental volume	35	1.00	1.04	1.11
svFMBV5	318	9.13	9.42	10.05
Composite of all triggers n (%)		969 (27.82)	863 (25.56)	651 (20.57)

4.3.3.3 CUSUM learning curve analysis

The CUSUM learning curve analysis for subgroup cohort 3 is shown in **Figure 17**, and the analyses for subgroup cohorts 1 and 2 are provided in **Appendix 1** and **Appendix 2**, respectively. Differences between the analyses in the three groups are negligible, and the analysis of cohort 3 is therefore the most relevant as it most closely approximates the intended real-world use of OxNNet.

Four of the 18 sonographers crossed the alarm line during the study (sonographers 2, 7, 12, and 18) and, other than those four, no sonographers crossed the warning line. Of those four, one (sonographer 18) improved spontaneously (that is, not in temporal relation to the delivery of the OxNNet feedback toolkit) and went on to significantly overshoot the reassurance line — a trend which continued for the next 350 scans. Two (sonographers 2 and 7) had crossed the alarm line at the time when performance was analysed and feedback delivered, and in both cases the OxNNet feedback toolkit agreed that QC had failed and urgent improvement was required. In the fourth case (sonographer 12), no analysis or feedback was performed because fewer than 50 scans had been undertaken at the time.

Of the three sonographers who received advisory comments on feedback (sonographers 4, 5, and 17), none crossed the warning line or the alarm line, and in fact all were close to or beyond the reassurance threshold at the time of feedback.

In five cases where the OxNNet feedback toolkit passed all QC checks (sonographers 1, 8, 9, 15 and 18), all had remained above the reassurance line for the preceding 50 scans.

Figure 17: Cumulative sum (CUSUM) learning curve analyses for each sonographer performance with respect to subgroup cohort 3 for the entire duration of the OxPLUS study

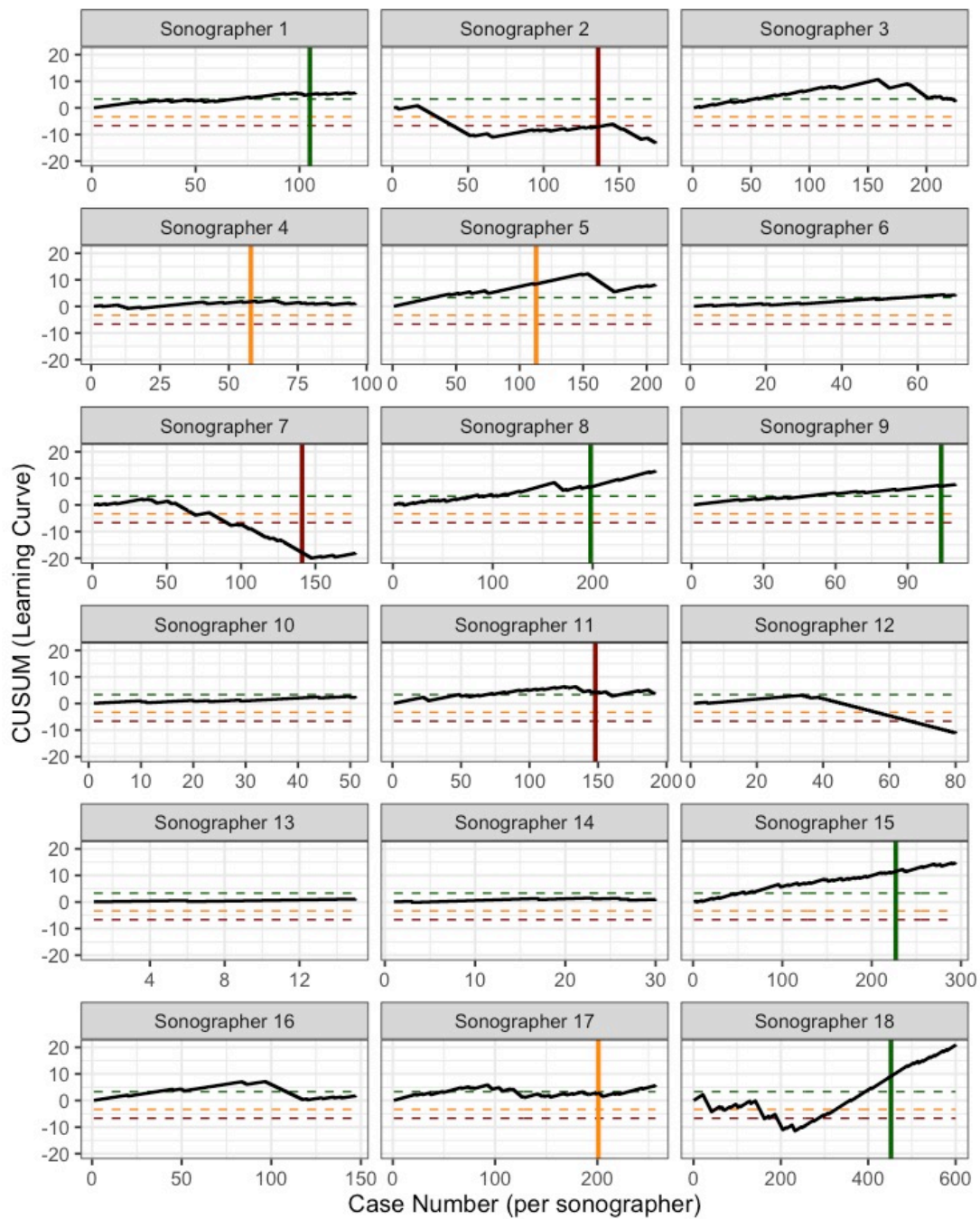
Incremental increases in the CUSUM curve represent improved performance.

Green dashed horizontal line (top): Reassurance threshold

Amber dashed horizontal line (middle): Warning threshold

Red dashed horizontal line (bottom): Alarm threshold

Solid vertical lines: The time at which the feedback was delivered using the OxNNet feedback toolkit; colour coded to represent pass (green), advisory only (amber), or failure in any domain (red).



4.4 Discussion and conclusion

This work demonstrates that it is possible to train a group of sonographers to perform first trimester 3D placental ultrasound with power Doppler, and that the initial learning curve is relatively fast. However, it is also evident that significant sources of potential bias exist, both with respect to measures of placental volume and svFMBV derived from those scans. The feasibility of a feedback toolkit has been demonstrated, and the potential added utility of multi-rule QC processes and CUSUM analysis has been shown.

4.4.1 Causes of QC failure

Measures of svFMBV appear particularly vulnerable to bias and QC failure. This likely reflects two interacting mechanisms: 1) problems with gain selection or signal calibration, and 2) problems with placental segmentation.

First, the FMBV measurement itself is technically sensitive: it requires careful attention to the region-of-interest, and consistent handling of sub-noise power Doppler gain setting selection, which may amplify small operator differences into large measurement variability — that is, bias in gain selection will likely result in bias of svFMBV measures. However, while it may be considered somewhat subjective, the selection of the sub-noise gain has been shown to have excellent reproducibility, demonstrating intraclass correlation coefficients of 0.98 (95% CI 0.97–0.99) and 0.98 (95% CI 0.98–0.99) inter-observer reproducibility between two pairs of expert assessors, and 0.96 (95% CI 0.93–0.98) for intra-observer reproducibility. This degree of reproducibility is quite remarkable for a subjective measure, and further work to verify these findings would be of benefit to understanding the limitations and sources of bias inherent to the svFMBV method. The single-vessel method of FMBV calibration (described in paragraph 1.4.2.1, page 34) should help to overcome inconsistencies in power Doppler gain settings and tissue attenuation.¹²⁴ In reviewing outlying results and QC failures, this process should also be scrutinised as a potential source of bias propagation, again an important subject for future work.

Secondly, svFMBV relies on accurate placental segmentation, and deviations in this process could contribute to QC failure. For instance, if the placenta lies heavily in shadow, the segmentation may not be precise, potentially placing the uteroplacental interface slightly within the myometrium or placenta

rather than at its actual location. This discrepancy can result in svFMBV being evaluated in regions with more or less vascularity than intended. Such errors are likely to affect the FMBV estimate, where even minor upstream mistakes could trigger significant downstream QC flags. Having developed a statistical model to identify QC failures in these areas, it would be valuable to visually review these cases (particularly with respect to placental segmentation) to better understand the factors contributing to outlier results and devise mitigation strategies — whether that involves improving sonographer training or refining OxNNet processes. These considerations should be addressed in future research.

4.4.2 Evaluation of OxNNet feedback toolkit and CUSUM learning curve

The results demonstrate that both the OxNNet feedback toolkit and CUSUM learning curve approach are capable of identifying patterns of systematic bias with respect to OxNNet-derived placental volume and svFMBV measures.

4.4.2.1 Feedback toolkit

A statistically significant change in the distribution of placental volume bias classifications was observed after delivery of the OxNNet feedback toolkit, but not with other parameters. The direction of that change was not uniform: the number of sonographers classified as passing rose from two to five, but the number classified as failing also rose from one to three, so that the distribution became more polarised rather than shifting consistently towards better performance. Bhapkar's test assesses marginal homogeneity and therefore indicates only that the distribution of classifications changed; it does not establish that performance improved. The uncontrolled before-and-after design, without a comparison group, cannot attribute any change to the feedback itself. This might simply reflect small case numbers — there were only nine sonographers in the paired comparison — and fewer evaluable images available for some sonographers in the post-feedback analysis. Other possible explanations include: 1) the OxNNet feedback tool in its current iteration may not be an effective means of delivering feedback; 2) there may be factors other than individual sonographer performance contributing to variation in bias and spread (such as bias in case selection where some sonographers are more likely to take on technically challenging cases, for example); and 3) the training delivered in response to feedback is not effective in correcting sonographer performance factors.

This is a novel tool, and as such there is no completely apposite literature available, but we can cautiously extrapolate from work on NT quality control. With respect to NT measurements, there is no evidence in the medical literature to suggest that this approach to audit and feedback is ineffective, and in fact there is substantial evidence to the contrary. In a study of 88 sonographers, Chalouhi et al. (2013) showed that those who received individualised feedback on bias of their serial NT measurements demonstrated marked improvements in their scoring. No significant change was observed among those who received no feedback. Importantly, the proportion of satisfactory images rose by nearly 50% among low scorers who were provided with feedback, underscoring the feasibility and impact of formative assessment in routine first-trimester practice.¹⁷⁸ Torrent et al. (2019) demonstrated that among 102 sonographers, periodic audits and feedback are associated with higher-quality NT measurements and reduced underestimation rates (24% vs 56%; $P = 0.021$).¹⁷⁹ However, evaluating 1.5 million NT scans in the Nuchal Translucency Quality Review (NTQR) program (United States of America), Cuckle et al. (2015) have shown that despite considerable training and audit, significant variability remains between NT providers.¹⁸⁰

An advantage of the feedback toolkit approach is that it does not require integration with OxNNet itself, or with any other digital systems. In real-world use, sonographers could collate and submit their own measurements at any time to receive feedback, and this could be made available in most settings as a Microsoft Excel file, as is the case for the NT self-assessment tool.^{168,169} This makes it readily deliverable and scalable across a range of different healthcare settings. Conversely, a major disadvantage is that this tool can only be applied in retrospect and requires sufficient cases for analysis. The more cases evaluated the more accurate the assessment is likely to become, but the longer the run of cases prior to feedback, the longer bias may go undetected.

4.4.2.2 CUSUM learning curve

The CUSUM learning curve method was applied retrospectively in this work but, with integration into the OxNNet software, CUSUM could provide running evaluation of individual sonographer performance including informing practitioners when they have crossed the warning or alarm thresholds.

There is consistent evidence that CUSUM-based prospective monitoring is a practical and sensitive method for maintaining and improving ultrasound measurement quality. Sabrià et al. (2011) applied a

CUSUM model to >2400 NT measurements performed by seven sonographers, and compared findings with the established Fetal Medicine Foundation (FMF) and the Women & Infants Hospital of Rhode Island (WIHRI) retrospective audit tools, demonstrating good agreement while allowing for earlier detection of measurement bias.¹⁸¹ Balsyte et al. (2011) showed, using a CUSUM chart of NT performance, that prompt identification of out-of-control performance and provision of corrective training resulted in improvement and return to in-control status.¹⁸² Sahota et al. (2012) applied CUSUM across four screening centres and 36 sonographers, demonstrating substantial variation between providers and identifying that nearly half of sonographers with ≥ 30 scans exhibited systematic under- or over-measurement, supporting routine prospective feedback to preserve integrity of screening.¹⁸³ CUSUM has also been used to assess QC and learning curves in various sonographic parameters, such as fetal biometry and anatomy screening, indicating its broader applicability across ultrasound measures for prospective audit.^{171,184}

A key drawback of the CUSUM approach is that to be used prospectively in practice, the CUSUM analysis tool must integrate with the OxNNet software, and this requires an accessible archive of every sonographer's prior scans. The version of OxNNet being prepared for market release is intentionally designed not to retain data, which offers clear information-governance benefits, but this makes integration of CUSUM analysis more challenging. Consequently, QC data — and potentially the CUSUM QC tool itself — would need to be hosted on local servers, adding technical and administrative complexity and rendering the solution impractical in settings that lack those resources or where sonographers frequently move between independent sites.

4.4.3 Strengths and limitations

This work is strengthened by the large number of cases evaluated, and the relatively large number of sonographers included in relation to other similar studies of sonographer performance measures. However, the lower numbers of sonographers and images available for evaluation after delivery of feedback are a significant limitation.

The availability of z-scores derived from peer reviewed and published reference charts is also advantageous, as is the use of robust statistical processes to determine the warning and alarm thresholds used for CUSUM learning curve analysis. However, basing this calculation on the frequency

of multi-rule QC failure within the same single dataset could lead to inflation or underestimation of the expected incidence of QC failure. This highlights the importance of ongoing calibration and adjustment of the tool as more data become available, and more is understood about the potential causes of outlying results.

This work is further limited by lack of literature reference points for comparison, which is inevitable to some degree when exploring a new marker. Extrapolation of learning from analyses of other sonographic parameters has some value but requires cautious interpretation.

4.4.4 Conclusion

Two separate tools have been created for auditing, providing feedback, and controlling the quality of metrics produced by OxNNet. Each tool offers unique advantages and disadvantages, delivering different benefits depending on the situation. While CUSUM is likely the more robust method for early detection of performance deviations, the feedback toolkit stands out for its accessibility and ease of use in real-world settings.

It is advisable to recalibrate both tools as additional data becomes available. This will help reduce the operational strain caused by false positives and decrease the likelihood that poor performance goes unnoticed.

[5] PREDICTION OF FETAL GROWTH RESTRICTION AND PRE-ECLAMPSIA

5.1 Introduction

The clinical significance of fetal growth restriction and pre-eclampsia has already been emphasised (paragraph 1.1, page 19), as has the importance of early prediction both for care planning and potential prophylaxis (paragraph 1.3, page 29).

For pre-eclampsia, robust multivariable models are now in clinical use, most notably the Fetal Medicine Foundation (FMF) competing-risks algorithm, which integrates maternal characteristics, mean arterial pressure, uterine artery Doppler, and placental growth factor. This approach, as implemented in the ASPRE trial, achieves high detection rates for preterm pre-eclampsia (approximately 75% at a 10% false-positive rate, and up to 82% in some validation cohorts) and has demonstrated a 62% reduction in preterm pre-eclampsia incidence with first-trimester aspirin prophylaxis. However, the predictive performance for term pre-eclampsia remains modest, with detection rates around 41%, and current models are less effective for prevention of term disease.^{86,185-}

¹⁸⁷ Systematic reviews confirm that the FMF model is the only externally validated tool with consistently high discrimination for preterm pre-eclampsia, while other models show variable and generally lower performance.¹⁸⁸

In contrast, prediction models for FGR are generally less mature. Most existing models predict small for gestational age (SGA) as a surrogate for FGR, and few have undergone robust external validation. Recent meta-analyses highlight that while some models, such as the International Prediction of Pregnancy Complications-FGR model, show promising discrimination and calibration, their clinical utility and impact on outcomes require further evaluation in routine practice.^{189,190}

A clear association between OxNNet-derived markers of placental volume z-score and svFMBV1-5 z-score has been demonstrated (paragraph 3.3.5, page 80) and they may have predictive utility. Numerous other biomarkers and factors from the medical history are known to have value in the prediction of PE^{86,187} and FGR⁸⁰, and therefore incorporation of OxNNet metrics into a multivariable

modelling approach is worthy of attention, particularly with respect to prediction of FGR and term PE, where there is a distinct clinical need for better predictive strategies.

5.1.1 Aims

- To develop multivariable predictive models for FGR and PE using the FirstPLUS dataset, incorporating OxNNet-derived 3D placental sonography markers.
- To externally validate those models using the OxPLUS dataset.

5.1.2 Statement of personal contribution

Early exploratory work was undertaken with the statistical guidance and supervision of Prof. Dave Wright, Professor of Medical Statistics, University of Exeter. The final statistical analysis plan was produced in discussion with Stephen Gerry, senior medical statistician, University of Oxford. The entire analysis was undertaken independently, with guidance sought as appropriate.

5.2 Methodology

5.2.1 Key datasets and definitions

This work utilises the FirstPLUS (n = 3479) and OxPLUS datasets (n = 2767), with exclusions as defined in paragraph 2.3 (page 52). The primary outcomes of interest are fetal growth restriction and pre-eclampsia.

5.2.2 Power calculation

Using Riley et al.'s (2019)¹³⁴ method via the *pmsampsize* package (v1.1.3)¹³³, calculations were performed using the actual outcome incidence and sample size from the FirstPLUS cohort, to determine the number of parameters that could be included in the final models, assuming constant shrinkage of 0.8 and a c-statistic of 0.7 or higher.

5.2.3 Exploration of collinearity

A correlation matrix was constructed to quantify associations among study variables. Continuous variables were analysed using Pearson correlations with pairwise deletion of missing data to maximise available information. The *cor* function in base R, and the *polychor* and *polyserial* functions from the *polycor* package, were used. Discrete variables (ordinal or nominal treated as ordered factors) were analysed using polychoric correlations to estimate the latent association between categorical responses. For mixed pairs (one continuous and one discrete), polyserial correlations were used, which model the continuous variable jointly with an underlying latent variable corresponding to the discrete measure. The resulting matrix was visualised as a heatmap. In this context, a latent variable is a continuous unobserved quantity that is assumed to generate the observed categorical responses.

5.2.4 Exploration of smoothing requirements

Univariable Cox proportional hazards models were fitted for each continuous candidate predictor variable, with fetal growth restriction (FGR) as the event and gestational age at birth as the time scale. Variables expressed as multiples of the median were $\log_{10}$ transformed prior to analysis; all other predictors were analysed on their original scale. To assess potential non-linearity, restricted cubic splines were fitted with 1, 2 and 3 degrees of freedom and the predicted log relative hazard (with 95%

confidence intervals) was plotted across the observed range of each predictor. The default placement of knots was used according to the rms package (10th, 50th and 90th percentiles for 2 degrees of freedom; and 5th, 35th, 65th, and 95th percentiles for 3 degrees of freedom).

These spline plots were inspected visually to determine whether a non-linear specification was warranted and, if so, the number of knots to retain, and AIC values were calculated and compared to confirm visual findings. The final degrees of freedom for each predictor were chosen to balance model parsimony and fit, considering both visual inspection and AIC values, favouring the simpler specification when additional knots produced minimal change in curve shape or suggested overfitting. Fewer degrees of freedom were preferred, since each additional degree of freedom will add a parameter to the final multivariable models.

This process was repeated for both the outcomes of FGR and PE, but also for normotensive FGR, any gestational hypertension (including PE), and a composite of any FGR or gestational hypertension.

5.2.5 Selection of variables for final models

The candidate variables were organised into categories: 1) sonographic markers, 2) serum biomarkers, and 3) factors from medical history along with mean arterial blood pressure. This classification is based on the practical consideration that measuring blood pressure and gathering information from medical history are usually feasible in most healthcare environments, whereas access to sonography and serum analysis may be limited. Therefore, it is helpful to treat these groups separately at this stage of the variable selection process.

For each variable group, a multivariable Cox proportional hazards model was fitted with either FGR or PE as the outcome variable, using the cph function from the rms package. Gestational age at birth was used as the time scale, and deliveries in which the outcome did not occur were treated as censored at that gestation. Because such deliveries are competing events rather than conventional censoring, these are cause-specific hazard models; the rationale for this choice, and its consequences for interpretation, are considered in the discussion. Relationships between predictor and outcome were treated as linear, unless otherwise specified, in which case restricted cubic splines were applied as described in paragraph 5.2.4 (page 116).

Each model was evaluated with ANOVA (using the `anova` function), and backwards selection was performed using the `fastbw` function in R. ANOVA plots were constructed and visually inspected, and output from the *fastbw* analysis (which produces a textual recommendation on variables for inclusion) was reviewed.

5.2.5.1 Final model construction

Final models were constructed for prediction of birth with FGR and PE by 42 weeks' gestation, and included all variables deemed to be statistically important via the backwards selection process. Aspirin use was not included in the backwards selection process but was 'forced' into all models.

Additional models were constructed using only the factors from the medical history, only the OxNNet metrics (placenta volume and svFMBV5), and both combined.

5.2.5.2 Handling of aspirin

Aspirin use among cases at high risk of PE in the training cohort was identified as a potential confounder of the relationship between predictors and outcomes, particularly PE. To control for this, aspirin use was forced into all multivariable models, so that the estimated effects of all other covariates are adjusted for both the potentially augmenting effect of being deemed high-risk enough to take aspirin, and the potentially diminishing effect of aspirin use reducing the outcome of interest. When applied to the OxPLUS external validation cohort, predictions were generated under a uniform "no-aspirin" scenario, so that every woman is scored on the same basis and the aspirin term contributes equally to every linear predictor. Because aspirin use here marks high-risk status rather than estimating a treatment effect, these predictions should not be read as the risk that would have been observed had aspirin not been taken.

A sensitivity analysis was undertaken to examine the consequences of this choice. Five scenarios were compared on external validation: predictions with the aspirin term removed from the linear predictor; predictions with aspirin use set to zero for every woman, as in the primary analysis; predictions using each woman's actual NICE-derived indication for prophylaxis; predictions restricted to women for whom prophylaxis was not indicated; and predictions restricted to women with an observed uterine artery pulsatility index, the last of these testing the effect of the median imputation described above. Discrimination was quantified by Harrell's C statistic and compared with the primary analysis in each case.

5.2.6 Validation of models

5.2.6.1 *Internal validation*

Apparent discrimination was quantified using the area under the time-dependent receiver operating characteristic curve (AUC) computed for 42 weeks' gestation, using the survivalROC package, with time-dependent sensitivity and specificity (and hence the AUC) estimated using Kaplan–Meier survival estimates for censoring (incomplete observations at the specified time) and outcome status. Bootstrapping was performed to generate 95% confidence intervals based on probabilities at the 2.5th and 97.5th quantiles.

Optimism-corrected AUC and 95% confidence intervals were obtained using a nonparametric bootstrap. For each replicate the data were resampled with replacement and the model refitted; AUC was computed both on the bootstrap sample (AUC_{boot}) and when that bootstrap-fitted model was applied to the original data (AUC_{test}). Replicate optimism was defined as ($AUC_{boot} - AUC_{test}$), and the replicate-level optimism-corrected AUC was calculated as the original apparent AUC minus this optimism. The final optimism-corrected AUC is the mean across replicates. A 95% bootstrap percentile confidence interval was formed from the 2.5th and 97.5th percentiles of the optimism-corrected AUC distribution.

5.2.6.2 *External validation*

For external validation the fitted models were applied to the OxPLUS cohort to generate subject-level risk scores and predicted survival probabilities at 42 weeks using `survest` and `predictSurvProb` in R, with individual risk of the outcome in question by 42 weeks calculated as (1 - survival probability).

Here, all cases were treated as FALSE with respect to aspirin use, irrespective of actual aspirin use. Where uterine artery Doppler PI measures were not available, these were treated as having values of 0, which corresponds to the population median because the variable is expressed in multiples of the median and $\log_{10}$ transformed. Imputation was necessary because the `cph` function cannot accommodate NA values in the dataset.

Uterine artery Doppler pulsatility index was not available for 660 of the 2746 women in the validation cohort (24.0%), whereas the measure was complete in the development cohort. The procedure described above therefore amounts to single imputation at the population median. External validation was performed in 2746 of the 2767 women who passed automated quality control, 21 being excluded because fetal growth restriction status could not be determined.

AUC was calculated from the C statistic and its standard error, extracted using the *rcorr.cens* function. A 95% confidence interval was then computed by the normal approximation ($C \pm 1.96 \times \text{standard error}$).

For each model, sensitivity, specificity, negative predictive value and positive predictive value were calculated at a 10, 30, and 50% false positive rate. ROC curves were constructed for each model.

This process was repeated for the following subgroups: 1) excluding cases of FGR and PE with delivery $\geq 37/40$, to evaluate the discrimination in preterm cases 2) exclusion of PE, to evaluate discrimination of the FGR prediction model for normotensive FGR, 3) excluding cases of pre-existing or gestational diabetes, to evaluate discrimination of the FGR prediction model without the potential confounder of diabetes-related fetal macrosomia.

5.2.6.3 Bootstrapping procedures

All bootstrap procedures reported here used 100 replicates to balance computational feasibility with timely analysis. For any resulting publications or work intended to inform clinical practice, increasing the number of bootstrap replicates to 1,000 is highly recommended to stabilise the percentile confidence intervals and improve the precision of the estimated sampling distribution. In these simulations, running 1,000 replicates substantially increases computational cost and wall-time, despite parallelising the bootstrap across multiple CPU cores. Therefore, 100 replicates were used as a

pragmatic compromise during development and exploratory analyses. A fixed random seed (`set.seed(123)`) was used to ensure reproducibility of the resampling.

5.2.7 Comparison of models

The DeLong test, a nonparametric method, was used to statistically compare AUC values, using the `roc.test` function from the `pROC` package in R. P values <0.05 were considered statistically significant.

5.2.8 Evaluation of model calibration

Calibration of each model was performed using the `calibrate` function from the `rms` package, with the time horizon set to 42 weeks ($u = 42$) and 100 bootstrap resamples ($B = 100$). Because this function resamples the data used to fit the model, calibration assessed in this way is internal to the FirstPLUS development cohort and is not an assessment in the external (OxPLUS) dataset. Agreement between predicted and observed probabilities was summarised by dividing the cohort into deciles of predicted risk and comparing the mean predicted risk within each decile with the Kaplan-Meier observed risk at 42 weeks. The mean absolute difference across deciles reflects overall calibration accuracy, while the 90th centile of that difference quantifies the magnitude of error in the worst-calibrated part of the range.

5.3 Results

5.3.1 Power calculation

With a 7.16% and 2.96% incidence of FGR and PE, respectively; assuming a c-statistic of 0.7 and shrinkage of 0.8, the 3479 cases in the FirstPLUS dataset should be sufficient to fit a model for FGR with 31 parameters if required, or a model for PE with 13 parameters.

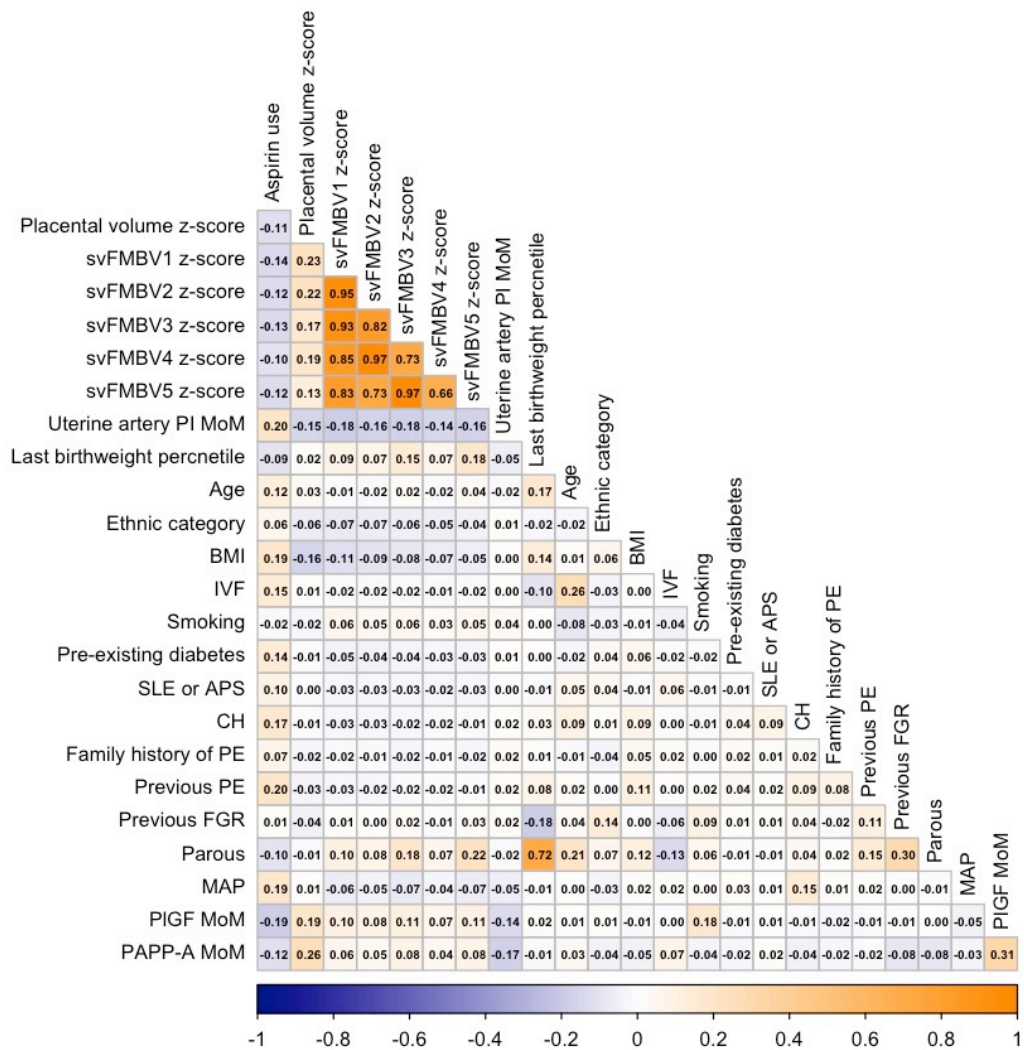
5.3.2 Correlation matrix

The correlation matrix is shown in **Figure 18** (page 124). The variables svFMBV1–5 are all highly correlated, and for that reason it was decided to select just one for further exploration, and svFMBV5 was chosen as this has the strongest associations with FGR and PE as shown in paragraph 3.3.5 (page 80).

The last birthweight percentile and parity were highly positively correlated (0.72), but this is expected because NA values for last birthweight percentile were treated as 0 for the purposes of this exercise, which corresponds to a code of 0 for nulliparity or 1 for parity. It was therefore deemed appropriate to take both forward for further consideration as candidate variables.

No other remaining variables demonstrated strong correlations and all were taken forward to the next round of variable selection.

Figure 18: Correlation plot showing associations between candidate variables



5.3.3 Exploration of smoothing requirements

The log relative hazard plots for placental volume z-score are presented in **Figure 19** (page 126) as an example, and all others are available in **Appendix 3** (page 209). The AIC values for each univariable model are provided in **Appendix 4** (page 216). It was decided to model maternal age, last birthweight percentile, and $\log_{10}$ uterine artery PI MoMs with two degrees of freedom (three knots) using restricted cubic splines for the prediction of both FGR and PE. In all other cases, a linear approach was chosen. The final decisions on degrees of freedom for each continuous variable are shown in **Table 19**.

Table 19: Decisions regarding degrees of freedom applied using restricted cubic splines to each continuous candidate variable

Variable	Degrees of freedom grouped by outcome of interest	
	FGR	PE
Maternal age	2	2
svFMBV5 z-score	1	1
Last birthweight percentile	2	2
$\log_{10}$ MAP MoM	1	1
$\log_{10}$ PAPP-A MoM	1	1
$\log_{10}$ PIGF MoM	1	1
$\log_{10}$ Uterine artery PI MoM	2	2
Placental volume z-score	1	1

Figure 19: Plots of placental volume z-score against log relative hazard (with 95% confidence intervals) for each outcome variable, exploring linear modelling and restricted cubic splines with 2 or 3 degrees of freedom.

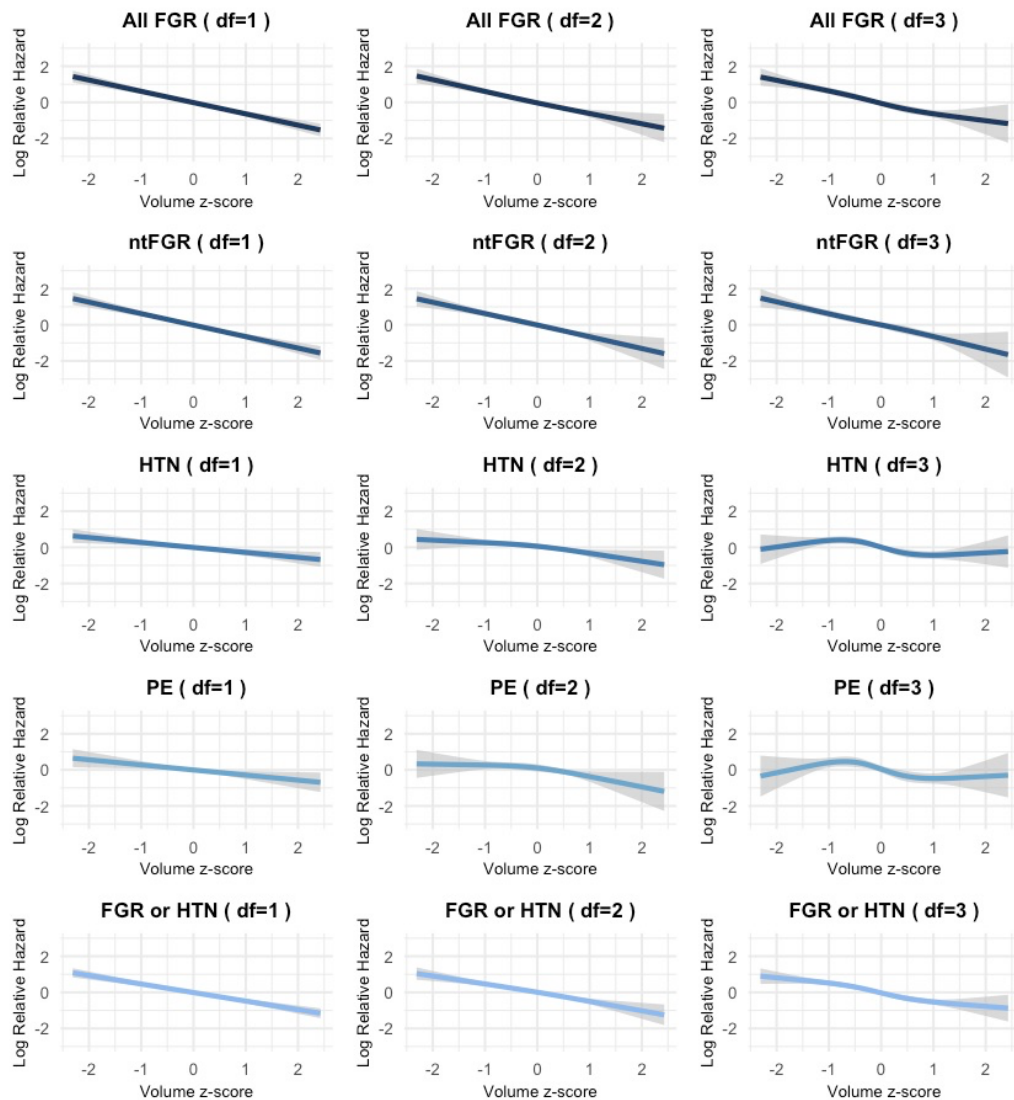
FGR: Fetal growth restriction

ntFGR: Normotensive fetal growth restriction

HTN: Any gestational hypertension, including PIH and PE

PE: Pre-eclampsia

FGR or HTN: A composite outcome of any FGR or HTN



5.3.4 Variable selection and model construction

The plotted ANOVA results for prediction of FGR and PE — including Chi-square and P values — are presented in **Appendix 5** and **Appendix 6**, respectively. The recommendations of the backwards selection process are shown in **Table 20**. On this basis, models were constructed for prediction of FGR and PE, as shown in **Table 21** (page 128). The non-linear element of $\log_{10}$ Uterine artery PI MoM demonstrated a statistically significant P value and therefore 2 degrees of freedom were preserved in the final models for prediction of both FGR and PE. This was also true for the non-linear elements of last birthweight percentile and maternal age, but these were only recommended for inclusion in the FGR prediction models and were excluded entirely from the PE prediction models.

Table 20: Recommendations for variables to include in predictive models based on ANOVA and backwards selection

Variable group	Variables	Recommendations	
		FGR	PE
Ultrasound markers	Volume z-score	Include	Exclude
	svFMBV5 z-score	Include	Include
	$\log_{10}$ Uterine artery PI MoM (non-linear)	Include	Include
Serum biomarkers	$\log_{10}$ PlGF MoM	Include	Include
	$\log_{10}$ PAPP-A MoM	Include	Exclude
History and MAP	Last BW% (non-linear)	Include	Exclude
	Age (non-linear)	Include	Exclude
	Race	Include	Exclude
	BMI	Exclude	Include
	IVF	Exclude	Exclude
	Smoking	Include	Exclude
	DM	Exclude	Include
	SLE APS	Exclude	Exclude
	CH	Include	Include
	FHPE	Exclude	Include
	Previous PE	Exclude	Include
	Previous FGR	Exclude	Include
	Parous	Include	Include
	$\log_{10}$ MAP MoM	Include	Include

Table 21: Final components of multivariable models for the prediction of fetal growth restriction and pre-eclampsia

Variables	FGR prediction models				PE prediction models			
	FGR1	FGR2	FGR3	FGR4	PE1	PE2	PE3	PE4
PV z-score	✓	✓		✓		✓		✓
svFMBV5 z-score	✓	✓		✓	✓	✓		✓
log ₁₀ Uta PI MoM	✓				✓			
log ₁₀ PIGF MoM	✓				✓			
log ₁₀ PAPP-A MoM	✓							
log ₁₀ MAP MoM	✓				✓			
Last BW% (non-linear)	✓		✓	✓				
Age (non-linear)	✓		✓	✓				
Ethnicity	✓		✓	✓				
Smoking	✓		✓	✓				
Chronic hypertension	✓		✓	✓	✓		✓	✓
Parous	✓		✓	✓	✓		✓	✓
BMI					✓		✓	✓
Diabetes mellitus					✓		✓	✓
Family history PE					✓		✓	✓
Previous PE					✓		✓	✓
Previous FGR					✓		✓	✓
Aspirin	✓	✓	✓	✓	✓	✓	✓	✓

5.3.5 Evaluation of model performance

The apparent AUC, optimism corrected AUC, and AUC derived from external validation for each model are shown in **Table 22**. The full models demonstrate apparent discrimination of 0.81 (95% CI 0.78–0.86) and 0.86 (95% CI 0.83–0.90) for FGR and PE, respectively. Modest reduction in discrimination was observed with optimism correction. Upon external validation, performance was further reduced, but remained good nevertheless, with external AUC of 0.75 (95% CI 0.66–0.84) for FGR and 0.80 (95% CI 0.72–0.88) for PE. Predictably, higher false positive rates were associated with greater sensitivity for the outcomes of interest, at the cost of lower specificity (**Table 23**). ROC curves for all models are presented in **Figure 20** (page 132).

The results of the aspirin sensitivity analysis were as follows. Because aspirin use enters each model as a single additive term, setting it to zero for every woman contributes the same quantity to every linear predictor and leaves the ranking of predicted risk unchanged; the C statistic under this scenario is therefore identical to that obtained with the aspirin term removed altogether, which was confirmed numerically. Substituting each woman's actual NICE-derived indication changed discrimination by between 0.000 and 0.073, in every case in the direction of better performance, indicating that the primary analysis does not overstate discrimination. Restricting validation to the 2326 women for whom prophylaxis was not indicated changed the C statistic by between –0.038 and +0.022, the models for FGR improving and those for PE deteriorating, the latter reflecting the substantial overlap between the NICE criteria and the predictors in those models. Restricting to the 2086 women with an observed uterine artery pulsatility index changed the C statistic by between –0.014 and +0.023; models that do not contain that variable changed by a similar amount, indicating that the difference reflects the characteristics of that subgroup rather than the imputation itself.

Table 22: Area under the ROC curve for each model including apparent and optimism corrected AUC from the training dataset, and AUC derived from external validation

Model name	AUC (95% confidence interval)		
	Apparent	Optimism corrected	External validation
FGR1	0.81 (0.78–0.86)	0.79 (0.76–0.84)	0.75 (0.66–0.84)
FGR2	0.73 (0.69–0.77)	0.73 (0.68–0.76)	0.62 (0.51–0.73)
FGR3	0.77 (0.74–0.80)	0.76 (0.72–0.80)	0.72 (0.62–0.81)
FGR4	0.81 (0.77–0.85)	0.80 (0.76–0.83)	0.74 (0.65–0.83)
PE1	0.86 (0.83–0.90)	0.85 (0.79–0.89)	0.80 (0.72–0.88)
PE2	0.64 (0.54–0.71)	0.64 (0.56–0.73)	0.63 (0.52–0.73)
PE3	0.76 (0.71–0.83)	0.76 (0.67–0.80)	0.69 (0.59–0.79)
PE4	0.76 (0.71–0.83)	0.75 (0.66–0.79)	0.71 (0.62–0.80)

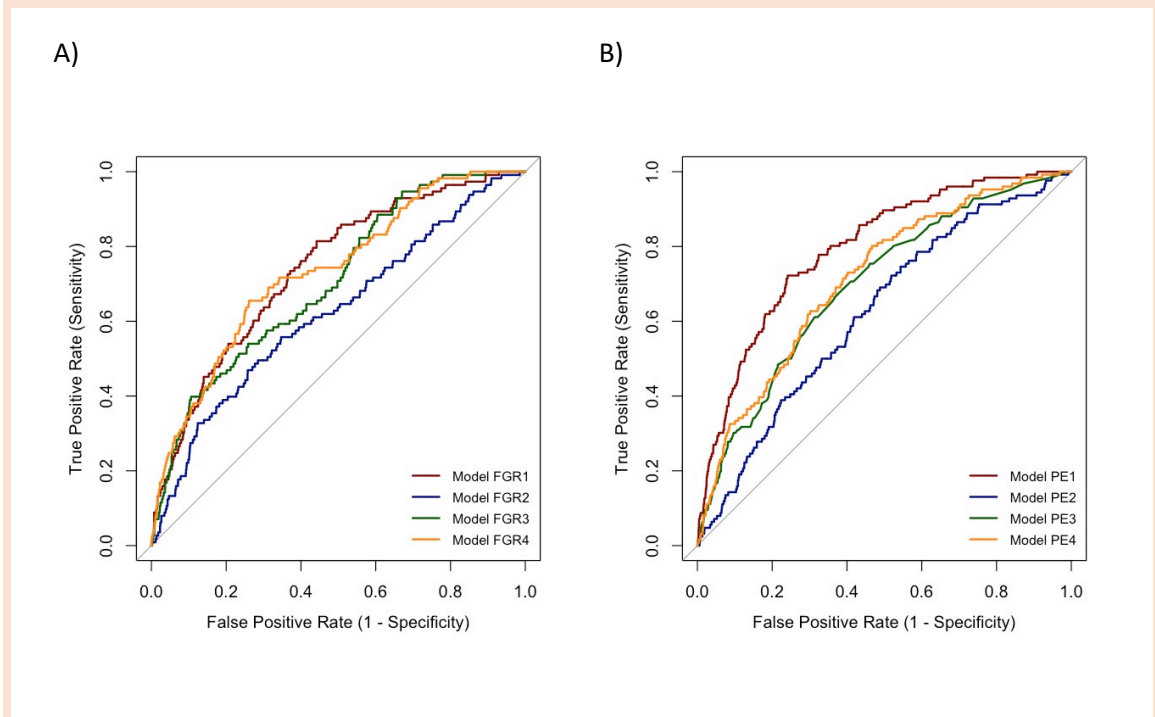
Table 23: Sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) for each model on external validation calculated for false positive rates (FPR) of 10%, 30%, and 50%

Model	10% FPR			
	Sens	Spec	NPV	PPV
FGR1	0.35 (0.27–0.45)	0.90 (0.88–0.91)	0.97 (0.96–0.98)	0.13 (0.09–0.17)
FGR2	0.25 (0.17–0.34)	0.90 (0.89–0.91)	0.97 (0.96–0.97)	0.09 (0.06–0.13)
FGR3	0.40 (0.31–0.49)	0.89 (0.88–0.90)	0.97 (0.96–0.98)	0.13 (0.10–0.17)
FGR4	0.38 (0.29–0.48)	0.89 (0.87–0.90)	0.97 (0.96–0.98)	0.13 (0.09–0.17)
PE1	0.44 (0.36–0.54)	0.89 (0.88–0.90)	0.97 (0.96–0.98)	0.16 (0.13–0.21)
PE2	0.17 (0.11–0.25)	0.89 (0.88–0.90)	0.96 (0.95–0.96)	0.07 (0.05–0.11)
PE3	0.32 (0.24–0.41)	0.88 (0.87–0.89)	0.96 (0.96–0.97)	0.11 (0.08–0.15)
PE4	0.34 (0.26–0.43)	0.88 (0.87–0.89)	0.97 (0.96–0.97)	0.12 (0.09–0.16)

Model	30% FPR			
	Sens	Spec	NPV	PPV
FGR1	0.64 (0.54–0.73)	0.70 (0.68–0.71)	0.98 (0.97–0.98)	0.08 (0.07–0.10)
FGR2	0.50 (0.40–0.59)	0.70 (0.68–0.72)	0.97 (0.96–0.98)	0.07 (0.05–0.09)
FGR3	0.56 (0.46–0.65)	0.70 (0.68–0.71)	0.97 (0.97–0.98)	0.07 (0.06–0.09)
FGR4	0.66 (0.57–0.75)	0.70 (0.68–0.71)	0.98 (0.97–0.99)	0.09 (0.07–0.11)
PE1	0.78 (0.70–0.85)	0.66 (0.65–0.68)	0.98 (0.98–0.99)	0.10 (0.08–0.12)
PE2	0.48 (0.39–0.57)	0.68 (0.66–0.70)	0.96 (0.95–0.97)	0.07 (0.05–0.09)
PE3	0.61 (0.52–0.70)	0.68 (0.66–0.70)	0.97 (0.96–0.98)	0.08 (0.07–0.10)
PE4	0.63 (0.54–0.71)	0.68 (0.66–0.70)	0.97 (0.97–0.98)	0.09 (0.07–0.11)

Model	50% FPR			
	Sens	Spec	NPV	PPV
FGR1	0.86 (0.78–0.92)	0.49 (0.47–0.51)	0.99 (0.98–0.99)	0.07 (0.06–0.08)
FGR2	0.63 (0.53–0.72)	0.50 (0.48–0.52)	0.97 (0.96–0.98)	0.05 (0.04–0.06)
FGR3	0.69 (0.60–0.77)	0.50 (0.48–0.52)	0.97 (0.96–0.98)	0.06 (0.04–0.07)
FGR4	0.74 (0.65–0.82)	0.51 (0.49–0.52)	0.98 (0.97–0.99)	0.06 (0.05–0.07)
PE1	0.90 (0.84–0.95)	0.47 (0.45–0.49)	0.99 (0.98–0.99)	0.08 (0.06–0.09)
PE2	0.72 (0.64–0.80)	0.48 (0.46–0.50)	0.97 (0.96–0.98)	0.06 (0.05–0.08)
PE3	0.79 (0.71–0.86)	0.48 (0.46–0.50)	0.98 (0.97–0.99)	0.07 (0.06–0.08)
PE4	0.83 (0.75–0.89)	0.48 (0.46–0.49)	0.98 (0.97–0.99)	0.07 (0.06–0.08)

Figure 20: Receiver operating characteristic (ROC) curves including all models for prediction of A) fetal growth restriction and B) pre-eclampsia



5.3.5.1 Comparison of basic care versus OxNNet

Pairwise comparison of model performance on external validation is shown in **Table 24** (page 133). The AUC is lower for FGR2 and PE2 (OxNNet alone) in comparison with the history only models, FGR3 and PE3 and this is statistically significant ($P = 0.021$ for FGR and $P = 0.024$ for PE). The combination of OxNNet metrics and medical history leads to a modest improvement over history alone for prediction of both FGR and PE, but this did not achieve statistical significance in either case.

Table 24: Pairwise comparison of AUC between models for prediction of fetal growth restriction and pre-eclampsia

First comparator		Second comparator		P
Model	AUC (95% CI)	Model	AUC (95% CI)	
FGR3 (History only)	0.72 (0.62–0.81)	FGR4 (History and OxNNet)	0.74 (0.65–0.83)	0.2
FGR3 (History only)		FGR2 (OxNNet only)	0.62 (0.51–0.73)	0.021
PE3 (History only)	0.69 (0.59–0.79)	PE4 (History and OxNNet)	0.71 (0.62–0.80)	0.062
PE3 (History only)		PE2 (OxNNet only)	0.63 (0.52–0.73)	0.024

5.3.5.2 Subgroup analyses

The AUC values for each model when applied to each of the subgroups are shown in **Table 25**. With term cases of FGR and PE excluded (subgroup 1) discrimination remained reasonably strong, and in some cases improved. Of note, the OxNNet variables on their own perform better in this subgroup with AUC improving from 0.62 (95% CI 0.51–0.73) to 0.68 (95% CI 0.48–0.88) for FGR, and 0.63 (95% CI 0.52–0.73) to 0.69 (95% CI 0.49–0.89) for PE, although the confidence intervals are wide and overlap substantially, reflecting the small number of preterm events in this subgroup. Strong performance was maintained by the full models (FGR1 and PE1) for prediction of both FGR and PE. When PE was excluded entirely, irrespective of gestational age at birth (subgroup 2), performance for detection of FGR was not materially altered (0.76 (95% CI 0.67–0.85) for the full model (FGR1), versus 0.75 (95% CI 0.66–0.84) in the full cohort). This indicates that discrimination for fetal growth restriction does not depend on the subset of cases in which it occurs alongside pre-eclampsia. In subgroup 3, which excluded all pre-existing and gestational diabetes, performance was similarly preserved, with no model differing from its full-cohort value by more than 0.01. The number of events within each subgroup is nonetheless small and the confidence intervals correspondingly wide, so these comparisons should be regarded as reassuring rather than definitive. The sensitivities and specificities for each model in each subgroup are available in **Appendix 7**.

Table 25: Performance of each model in subgroups for external validation

Subgroup 1: excluding cases of FGR and PE with delivery $\geq 37/40$

Subgroup 2: excluding PE

Subgroup 3: excluding pre-existing or gestational diabetes

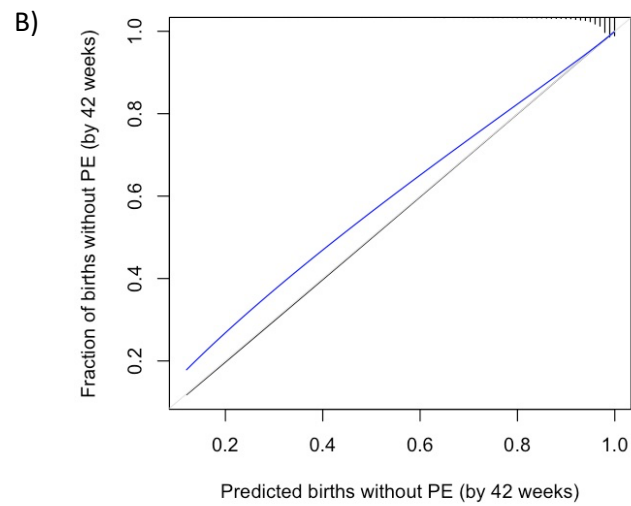
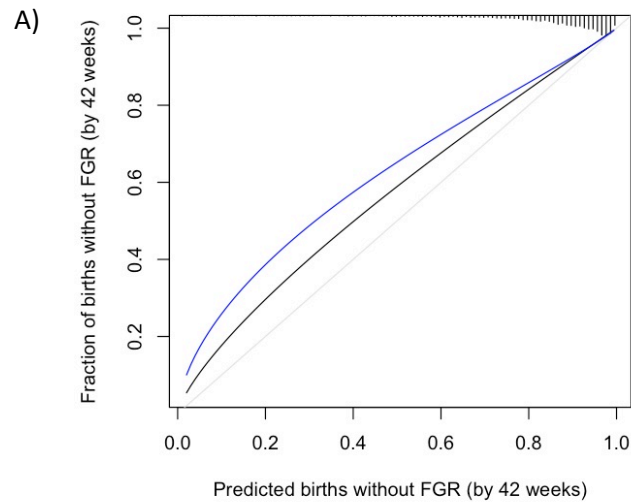
Model name	Externally derived AUC (95% confidence interval)			
	Full cohort	Subgroup 1	Subgroup 2	Subgroup 3
FGR1	0.75 (0.66–0.84)	0.73 (0.56–0.89)	0.76 (0.67–0.85)	0.75 (0.66–0.84)
FGR2	0.62 (0.51–0.73)	0.68 (0.48–0.88)	0.62 (0.50–0.74)	0.61 (0.50–0.73)
FGR3	0.72 (0.62–0.81)	0.64 (0.47–0.81)	0.74 (0.64–0.83)	0.73 (0.63–0.82)
FGR4	0.74 (0.65–0.83)	0.72 (0.56–0.89)	0.75 (0.65–0.85)	0.74 (0.64–0.84)
PE1	0.80 (0.72–0.88)	0.82 (0.64–1.00)		0.80 (0.71–0.89)
PE2	0.63 (0.52–0.73)	0.69 (0.49–0.89)		0.63 (0.53–0.74)
PE3	0.69 (0.59–0.79)	0.65 (0.41–0.89)		0.68 (0.57–0.79)
PE4	0.71 (0.62–0.80)	0.70 (0.49–0.91)		0.71 (0.60–0.81)

5.3.6 Evaluation of model calibration

Calibration analyses demonstrated good agreement between predicted and observed probabilities for both complete models, FGR1 and PE1, within the FirstPLUS development cohort. For the full FGR prediction model (FGR1), the mean absolute difference between predicted and observed risk across deciles of predicted risk was 0.034, indicating that on average predicted risks deviated from observed risks by approximately 3%. The 90th centile of that difference was 0.057, so that even in the worst-calibrated decile of predictions the discrepancy remained below six percentage points. For the full PE prediction model (PE1), agreement was closer still, with a mean absolute difference of 0.015 (1.5%) and a 90th centile of 0.029. These results indicate that both models were well calibrated at 42 weeks within the cohort in which they were developed, with PE1 showing particularly close alignment between predicted and observed outcomes. Because these are apparent estimates, obtained in the same cohort that was used to fit the models, they are optimistic by construction and do not describe calibration in an independent population. The calibration plots for each full model are presented in **Figure 21** (page 137).

Figure 21: Calibration plots for A) the full FGR prediction model (FGR1) and B) the full PE prediction model (PE1)

Grey line: Ideal calibration
Black curve: Observed calibration
Blue curve: Optimism corrected calibration



5.4 Discussion and conclusion

This study has shown that multivariable predictive models for FGR and PE — including OxNNet-derived svFMBV5 and, for fetal growth restriction, first trimester placental volume, along with other key biomarkers and medical history factors — offer good discrimination for both conditions. Discrimination was attenuated but preserved after optimism correction and on assessment in an independent cohort, following a rigorous process of variable assessment and backward selection.

A model predicting FGR using only OxNNet-derived first trimester placental volume and svFMBV5 reached moderate accuracy on external validation, and the corresponding model for PE performed similarly, with external AUC values of 0.62 and 0.63 respectively. These findings indicate that OxNNet-derived metrics might merit evaluation as a low-cost triage tool where other biomarkers or detailed medical histories are unavailable (such as in resource-limited environments), although on external validation both models discriminated significantly less well than the corresponding models based on medical history alone, and detected only 17–25% of cases at a 10% false positive rate. This is of interest because OxNNet could eventually be integrated into a bedside device with automated processes and built-in quality control, allowing use with minimal training when full ultrasound assessments are not practical. Although this idea is currently speculative, the results suggest it may be worth pursuing.

Subgroup analysis revealed a slight decrease in the full model's performance for predicting FGR when cases of term FGR were excluded. Conversely, the AUC for predicting PE improved modestly when term cases were omitted. Excluding all cases of PE did not materially alter model performance for predicting FGR (i.e., FGR without concurrent PE), suggesting that discrimination is not dependent on the subset of cases in which the two conditions occur together. Performance was likewise maintained when pregnancies complicated by pre-existing or gestational diabetes were excluded. This is reassuring, since diabetes can confound the assessment of growth by increasing fetal growth velocity and potentially masking FGR diagnosis, so that their exclusion might have been expected to improve model performance. One possible explanation for the absence of any such change is collinearity between gestational diabetes and ethnicity — known to be an important risk factor for gestational diabetes¹⁹¹ — which had a significant impact on the final model. A case-by-case exploration of those cases with concurrent gestational diabetes and FGR could prove insightful, and this should be the subject of further work.

5.4.1 Existing literature

5.4.1.1 *Prediction of fetal growth restriction*

Numerous attempts have been made to construct predictive models for fetal growth restriction (FGR) with some success, although many rely on small for gestational age (SGA) as a proxy outcome. However, few of these models have undergone rigorous external validation. In 2024, Allotey and colleagues conducted one of the largest studies to date on FGR prediction, aiming to address this gap by utilising the International Prediction of Pregnancy Complications (IPPIC) Network database. This repository contains harmonised individual-level data from large birth cohorts, registries, and hospital records, enabling an extensive meta-analysis of 55 studies encompassing 119 predictive models. Their focus was on models predicting FGR as a binary outcome or birthweight as a continuous variable.¹⁸⁹

A key methodological feature of this study was the definition of FGR, which was specified as birthweight below the 10th centile in combination with serious adverse outcomes such as stillbirth, neonatal death, or delivery before 32 weeks' gestation. This differs from the definitions employed in the FirstPLUS and OxPLUS studies. The authors justified their approach by excluding constitutionally small but healthy infants and by incorporating early delivery to mitigate treatment paradox. While this definition likely captures the most severe cases of FGR, it excludes infants born at ≥ 32 weeks who survived, potentially thanks to identification of FGR and obstetric intervention, and this is a group that remains clinically important. The authors also highlighted substantial heterogeneity in FGR definitions across the literature, noting that none of the existing models aligned precisely with their chosen outcome. As a result, despite the stated aims of the project, none of the previously published models predicting FGR as a binary event could be externally validated.

The authors therefore developed an entirely novel model, the IPPIC-FGR model, using pooled data from four IPPIC cohorts (237,228 pregnancies). The model was constructed using random intercept regression with backward elimination for variable selection, incorporating predictors such as maternal age, height, parity, smoking status, ethnicity, history of hypertension, and prior adverse pregnancy outcomes (pre-eclampsia, stillbirth, or SGA). Gestational age at delivery was also included, allowing risk estimation at specified gestational ages by imputing an assumed delivery date. While this approach yielded an apparent c-statistic of 0.96 (95% CI 0.51–1.0), suggesting excellent discrimination, performance declined markedly when applied across all women irrespective of gestational age at

delivery, with the pooled c-statistic falling to 0.66 (95% CI 0.26–0.91). This raises concern that inclusion of gestational age at delivery accounts for a substantial portion of this model's predictive power. This may introduce bias, given that gestational age at birth also forms part of the study's definition of FGR. That is, the definition of FGR as birth <32 weeks' gestation with birthweight <10th centile is far more likely to be fulfilled where the gestational age at birth is assumed to be < 32 weeks, whereas this aspect of the definition cannot be fulfilled if the gestational age at birth is >32 weeks. Arguably, this is a major limitation.

Finally, although the study drew on four large datasets, some cohorts were relatively small or had low event rates, limiting the ability to externally validate the IPPIC-FGR model. Thus, even this model, which was created as part of a study with external validation in mind, is not ready for clinical use.

Another notable example is the Fetal Medicine Foundation (FMF) competing-risks model, which incorporates maternal characteristics, uterine artery pulsatility index (UtA-PI), and biochemical markers (PAPP-A, PlGF).¹⁹² This model has a reported area under the curve of 0.82 for predicting birthweight below the 10th centile at <37 weeks' gestation, although confidence intervals were not provided. Subsequent robust external validation in a large European cohort demonstrated an AUC of 0.79 (95% CI 0.76–0.82), which still demonstrates good discrimination, but this highlights the importance of external validation to understand the true discriminatory capacity of a novel model.¹⁹³ The competing-risks framework offers the distinct advantage of being able to predict any birthweight z-score at any gestational age. However, its reliance on a fixed centile cut-off for birthweight (albeit one that can be manually adjusted) limits its scope, as it does not capture other clinically relevant features of FGR such as adverse perinatal outcomes, reduced growth velocity, or abnormal Doppler findings. Essentially, this is a predictive model for SGA, not FGR, and as such its clinical utility is limited.

5.4.1.2 Prediction of pre-eclampsia

Multivariable models for the prediction of PE are generally at a greater level of maturity compared with prediction of FGR, and there is a wealth of first trimester prediction models available. The Fetal Medicine Foundation competing risks model for PE prediction is of particular note because of its predictive performance which withstands external validation, and because it has been incorporated into the 'Tommy's app' — a clinical decision support tool being introduced in UK maternity care to improve risk assessment for complications such as pre-eclampsia.^{66,194}

The FMF competing-risks model for PE combines maternal factors, mean arterial pressure, uterine artery pulsatility index, and serum biomarkers (PIGF and PAPP-A). Large multicentre validation studies have confirmed the effectiveness of this approach. O’Gorman *et al.* (2017)¹⁹⁵ demonstrated high detection rates for very early, preterm, and term pre-eclampsia in a European cohort, while a UK NIHR study found that the FMF approach significantly outperformed the NICE method of PE risk assessment, achieving an 82% detection rate for preterm pre-eclampsia.⁸⁷ Further validation in Asia confirmed comparable performance across different populations, marking a landmark in external validation of this approach.¹⁹⁶ Secondary analysis of the ASPRE trial reinforced the value of risk-based screening using maternal factors, mean arterial pressure (MAP), uterine artery pulsatility index (UtA-PI), and biomarkers such as PAPP-A and PIGF, further demonstrating that this approach stratifies risk more effectively than categorical NICE or ACOG criteria, reducing false positives and aligning incidence rates with background risk.¹⁹⁶ Collectively, these studies demonstrate that biomarker-based FMF models provide accurate, reproducible, and clinically useful prediction of pre-eclampsia across diverse populations. Discrimination for preterm pre-eclampsia has been consistently shown to be excellent, with pooled area under the curve (AUC) values of 0.90 (95% prediction interval 0.76–0.96) across 16 external validation studies¹⁸⁸, and AUCs of 0.87 (95% CI 0.79–0.92) in large Australian cohorts.¹⁹⁷ Detection rates for preterm pre-eclampsia at a 10% false-positive rate are typically 70–75%.^{188,197,198}

5.4.1.3 Models including placental volume and fractional moving blood volume

This is not the first study to explore first trimester placental volume for prediction of FGR. Papastefanou *et al.* (2018), in a cohort of 3358 singleton pregnancies, demonstrated that a multivariable model including placental volume z-score, maternal weight, and uterine artery pulsatility index achieved an AUC of 0.87 for FGR and 0.92 for PE.¹⁹⁹ This study used a similar definition of FGR to OxPLUS and FirstPLUS, including birthweight <10th centile with Doppler abnormalities, although birthweight <3rd centile without Doppler abnormalities was not included.

While the results are promising and comparable to our own, no external validation was performed, or indeed optimism correction. Placental segmentation and volume calculation was performed using VOCAL™, and z-scores for placental volume were computed by applying the anti-log₁₀ function to the predicted values derived from the regression of log₁₀ PV against CRL.¹⁹⁹ In comparison, the work

presented in this chapter is strengthened by the use of OxNNet for segmentation and volume assessment, and the use of formal reference charts for the derivation of placental volume z-scores.

González-González et al. (2017), in a cohort of 988 singleton pregnancies, demonstrated that a multivariable model including placental volume and three-dimensional power Doppler vascular indices (vascularisation index, flow index, and vascularisation flow index), alongside maternal characteristics and uterine artery pulsatility index, achieved an area under the curve of 0.74 (95% CI 0.70–0.77) for FGR, defined as birthweight <10th centile by customised birthweight curves.²⁰⁰ This study is therefore limited by its definition of FGR, which does not incorporate Doppler abnormalities or other adverse outcomes. The addition of placental volume and vascular indices significantly improved model performance compared to models based only on maternal characteristics and first-trimester screening variables, but the sensitivity remained too low for clinical utility. This study also used VOCAL™ for placental segmentation and volume calculation, and vascular indices were derived from three-dimensional power Doppler ultrasound. The use of these vascular metrics is an additional weakness, since they are unstandardised and highly influenced by power Doppler gain settings (as discussed in paragraph 1.4.2, page 34). The use of svFMBV in OxPLUS and FirstPLUS is therefore a comparative strength. Further, although 10-fold cross validation was undertaken (which helps to account for optimism), no external validation was performed.²⁰⁰

5.4.2 Strengths and limitations

In contrast to the studies of Papastefanou et al. (2018)¹⁹⁹ and González-González et al. (2017)²⁰⁰, this work is strengthened by its comparatively large sample size. Indeed, it represents the largest cohort reported to date investigating first-trimester placental metrics in relation to adverse pregnancy outcomes. The availability of two distinct datasets, one suited to model development and the other to external validation, further enhances the robustness of the findings. Nevertheless, when compared with studies that have produced the most successful predictive models for FGR and PE, the sample remains relatively modest. Although power calculations support the adequacy of the data for model fitting, the low number of preterm events may limit model calibration and performance at earlier gestations. This is particularly important given that preterm FGR and PE are typically the most severe and clinically relevant cases, and this should be considered a limitation.

A further strength of this work lies in the definition of FGR, which extends beyond an arbitrary birthweight centile threshold to incorporate Doppler abnormalities, thereby providing a more clinically meaningful outcome than birthweight-based SGA alone. For PE, the use of a definition that does not conflate the condition with FGR allows clearer delineation of the two disease processes, and this aligns with the definition of PE used for the development of the FMF competing risks models which makes comparison more valid.

5.4.2.1 Modelling approach

The decision to employ the Cox proportional hazards approach is also a relative strength. By treating birth without FGR or PE as survival outcomes, it is possible to adjust predictions to evaluate the risk of birth with the condition in question at any given gestational age. This approach overcomes the potential bias introduced by forcing an estimate of the unknown gestational age at birth into the model, but the performance of the models at key preterm gestations (such as <32 weeks' gestation) has not been explored. This would be a useful focus of future work.

It is important to be explicit about what this framework estimates and what it assumes. Both FGR and PE are ascertained at delivery rather than at an observable point during pregnancy, and delivery without the outcome is not censoring in the conventional sense but a competing event: once a woman has delivered an unaffected infant she is no longer at risk. Because gestational age at birth is the time scale, every observation contributes its delivery time and there is essentially no administrative censoring, so that all non-events are competing events. Treating those competing events as censored is what defines the cause-specific hazard, and the models fitted here are therefore cause-specific hazard models. The quantity estimated is the instantaneous rate of delivering with FGR or PE at a given gestational age among women still pregnant at that gestation. This is well defined and estimable from the observed data, and it requires no assumption of a latent onset time preceding delivery; the model describes the gestational age distribution of affected births relative to unaffected ones. The same structure underlies the Fetal Medicine Foundation competing-risks models, in which the distribution of gestational age at delivery with pre-eclampsia is modelled with delivery without pre-eclampsia as the competing event.

Two consequences follow. First, discrimination is unaffected by the competing-risks structure, because the area under the curve, and the sensitivity and specificity derived from it, depend only on the

ordering of the linear predictor, and that ordering is valid under a cause-specific formulation. Second, survival probabilities derived from these models are net rather than crude probabilities: they describe the risk that would obtain if delivery without the outcome could not intervene, and they therefore exceed the corresponding cumulative incidence. Where calibrated absolute risks are the objective, a subdistribution hazard model of the Fine–Gray type, or an Aalen–Johansen estimator of the cumulative incidence function, would be the appropriate choice. The purpose here was to rank women by risk rather than to assign absolute probabilities, and the cause-specific formulation was considered appropriate on that basis, but the distinction should be borne in mind whenever the predicted probabilities themselves are read.

5.4.2.2 *Handling of aspirin*

The handling of aspirin presented a particular methodological challenge, given its potential to introduce treatment paradox where preventive interventions reduce the incidence of the outcome of interest and thereby diminish apparent predictive accuracy. In this study, aspirin use was forced into the models but all cases were treated as non-aspirin users for external validation. While this approach likely results in overly optimistic apparent discrimination, it was considered the most appropriate option. This is because aspirin use in the FirstPLUS cohort was recommended on the basis of the FMF competing risks assessment for PE, meaning that aspirin exposure effectively identified individuals already deemed high risk by factors included in the final FGR and PE models. As a result, risk assessment implicitly encoded high-risk status, potentially shrinking other coefficients and inflating apparent and optimism-corrected discrimination, while diminishing estimates on external validation. Nonetheless, this strategy ensures that every woman is scored on a common basis, and external validation suggests that the reported performance is, if anything, a slight underestimate.

One further consideration should be acknowledged. The fitted coefficient for aspirin was positive in every model, corresponding to hazard ratios of between 1.16 and 5.67, so that aspirin use marked increased rather than reduced risk. This reflects confounding by indication rather than any effect of treatment, since aspirin was recommended to women already identified as being at high risk. The pattern across models supports that interpretation, the coefficient attenuating as other risk factors are added, from a hazard ratio of 2.32 in the model containing only OxNNet-derived metrics to 1.16, and no longer statistically significant, in the full model for FGR. It follows that setting aspirin use to

zero at validation removes a marker of risk rather than an estimate of treatment benefit, and the resulting predictions should not be read as the risk that would have been observed had aspirin not been taken. This does not affect the reported discrimination, which the aspirin term cannot influence, and the sensitivity analysis indicates that the approach taken is conservative.

Alternative strategies were considered but ultimately rejected. Restricting model fitting to women not recommended aspirin would have excluded many of the highest-risk cases, disproportionately reducing event numbers for PE and introducing systematic bias by omitting extreme values of continuous predictors. A second option, modelled on the approach of Wright et al. (2022), involved imputing prevented PE cases using Markov chain Monte Carlo methods and pooled estimates based on biomarker-driven logistic regression, assuming a fixed aspirin effect size from ASPRE.²⁰¹ While methodologically sophisticated, this approach risked bias by using the same predictors for both imputation and model construction, potentially strengthening associations artificially and contributing to overfitting. Moreover, it relied on a pre-specified probability of aspirin efficacy, which is well established for PE but less certain for FGR given the heterogeneity of definitions in the literature. For these reasons, both alternatives were rejected.

5.4.2.3 Bootstrapping limitations

It is standard practice to perform bootstrap resampling with at least 1000 replicates for analyses intended for publication, particularly when results are to be applied in a clinical context. This level of resampling stabilises confidence interval estimation and enhances the precision of performance metrics. However, the computational cost of generating 1000 bootstrap samples is considerable and can be highly time-consuming. Although parallelisation — engaging all available processor cores to run resampling iterations concurrently — can reduce runtime, analyses of this scale remain computationally intensive, often requiring overnight execution for each full bootstrap cycle. The results presented here should therefore be regarded as preliminary. While increasing the number of replicates to 1000 would be expected to have only a negligible impact on point estimates, this step should nonetheless be undertaken before the findings are advanced further. Future work will focus on optimisation and streamlining of the R code to maximise computational efficiency, alongside the use of high-performance computing resources to accommodate the demands of large-scale resampling.

5.4.3 Conclusion

In summary, this work demonstrates that multivariable predictive models incorporating OxNNet-derived placental metrics alongside established biomarkers and maternal risk factors can achieve good discrimination for both FGR and PE, which was attenuated but preserved after optimism correction and on assessment in an independent cohort. The methodological choices — including rigorous variable selection, use of survival modelling, and careful handling of aspirin exposure — enhance the credibility of the findings, while the large cohort and availability of distinct datasets provide a solid foundation for future refinement. Nonetheless, limitations such as low preterm event rates, computational constraints on resampling, and the need for further exploration of subgroup effects highlight areas for ongoing development. These results support the potential clinical utility of OxNNet-derived placental measures and provide the basis for future work aimed at optimising prediction of the most severe and clinically relevant cases of FGR and PE.

[6] DEVELOPMENT OF NOVEL MARKERS

6.1 Introduction

When performing a three-dimensional (3D) first-trimester ultrasound with power Doppler, the potential for extraction of relevant ultrasound markers extends beyond placental volume and fractional moving blood volume. Additional morphological and vascular markers can be derived, some of which may hold predictive value for fetal growth restriction (FGR) and pre-eclampsia (PE).

In addition to placental volume, it has been suggested that the placental surface area may have an association with FGR, and placental thickness has also been explored as a possible marker of both FGR and PE.^{202,203} This concept has face validity because the same placental volume can arise from markedly different morphologies — a spherical, compact placenta would yield a relatively small uteroplacental interface (UPI), whereas a thin, flattened placenta of equal volume would present a substantially larger UPI surface area.

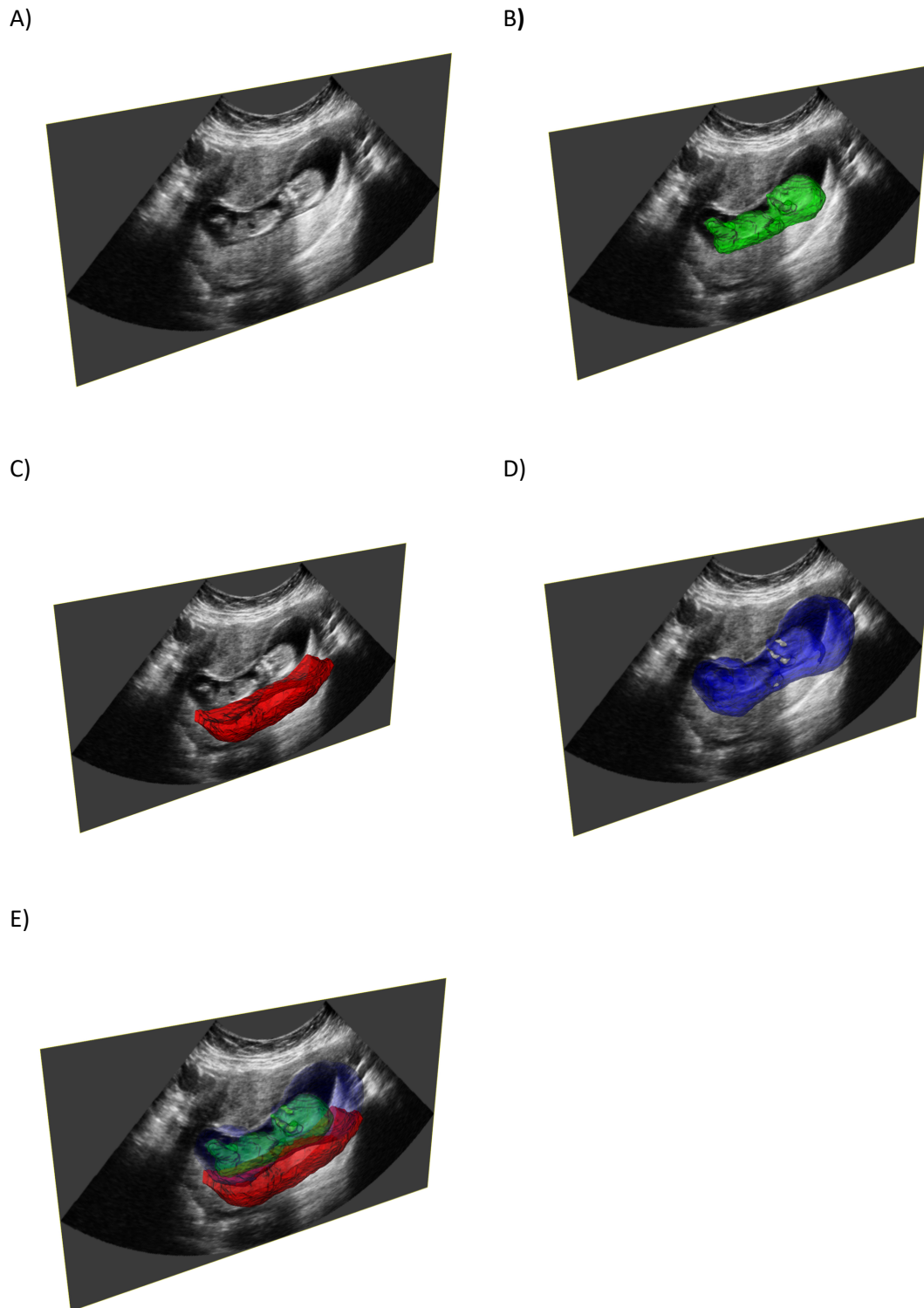
OxNNet, in addition to enabling placental segmentation, is capable of multiclass segmentation, allowing simultaneous delineation of the placenta, fetus, and amniotic fluid (**Figure 22**, page 149).¹¹² The uteroplacental interface can be defined as that portion of the placental surface not in contact with amniotic fluid or fetal structures, corresponding anatomically to the basal plate where maternal tissues meet the placenta. This boundary provides a reproducible anatomical reference point, thereby offering a robust anchor for further morphological characterisation.

Advances in three-dimensional placental imaging, particularly through the integration of power Doppler, now enable direct interrogation of the uteroplacental vasculature in vivo, extending beyond the established assessment of fractional moving blood volume. By the late first trimester — when spiral artery remodelling should be in progress if not largely complete⁴³⁻⁴⁸ — the placenta and its vascular connections can be visualised in their entirety. By employing artificial intelligence, these imaging modalities allow vascular features to be automatically extracted and systematically analysed, opening new opportunities for early evaluation of placental function and its association with adverse pregnancy outcomes. Of particular interest is the characterisation of spiral artery jets, given that both histological and sonographic changes in these vessels have been documented. Their successful

interrogation may in time contribute evidence relevant to the proposed 'jet-hose' hypothesis in the pathogenesis of pre-eclampsia and possibly FGR.^{61,62} Moreover, detailed assessment of radial and arcuate vessel characteristics could yield complementary insights, bearing on or refuting modelling studies that suggest upstream uterine artery changes cannot be explained by failed spiral artery remodelling alone, but may also require alterations in radial and arcuate vessels.⁶⁹

The placenta and its vasculature may therefore harbour numerous novel, so far unexplored markers with the potential to enhance prediction of FGR and PE, while also deepening our understanding of the underlying pathophysiology of these conditions.

Figure 22: A two-dimensional B-mode ultrasound image of a first trimester pregnancy with three-dimensional OxNNet-derived multiclass segmentation masks overlaid, showing A) the plain two-dimensional ultrasound, B) fetal segmentation, C) placental segmentation, D) amniotic fluid segmentation, and E) complete multiclass segmentation



6.1.1 Aims

- To develop and implement a clinically driven strategy to extract novel morphological and vascular markers from the extensive bank of 3D placental scans generated by the OxPLUS and FirstPLUS studies.
- To investigate, on an exploratory basis, the associations between these newly derived markers and key adverse pregnancy outcomes, FGR and PE.
- To explore how the findings fit with our understanding of the pathophysiology of placental insufficiency syndromes.

6.1.2 Statement of personal contribution

This work was undertaken in collaboration with Dr. Mohammadreza Soltaninejad, computer scientist in ultrasound image analysis at the Nuffield Department of Women's and Reproductive Health, University of Oxford. The design and execution of all machine learning and computational processes required to extract novel metrics were performed by Dr. Soltaninejad. The strategy to interrogate the uteroplacental vasculature was developed in collaboration with Dr. Soltaninejad who provided the technical expertise while I provided clinical and anatomical guidance. The analysis and interpretation of the resulting data are entirely my own work.

6.2 Methodology

6.2.1 Sources of data

This work utilises both the FirstPLUS (n = 3479) and OxPLUS (n = 2767) cohorts, combined into a single dataset (n = 6246). The derivations of the FirstPLUS and OxPLUS cohorts are described in detail in paragraph 2.3 (page 52).

6.2.2 Image processing

6.2.2.1 Uteroplacental interface surface area

An established segmentation framework was employed to classify the placenta, fetus, and amniotic fluid in three-dimensional B-mode (greyscale) ultrasound images using OxNNet.¹¹² As with many automated segmentation techniques, the initial outputs can include errors such as false positives and false negatives, often manifesting as discontinuities or small voids that do not correspond to the true continuity of tissue structures. To mitigate these artefacts, a post-processing ‘smoothing’ procedure was developed to reduce noise and surface irregularities. An artefact of this process is the potential to create new discontinuities (or ‘holes’), in which case the largest connected mass is identified and gaps filled using morphological operations, which assumes that the area of interest is a single mass.

From these outputs, the uteroplacental interface (UPI) was identified, defined as the portion of the placental surface directly adjacent to the uterine wall.²⁰⁴ This was achieved by excluding placental regions in contact with the areas labelled as amniotic fluid or fetus. To ensure precision, the UPI boundary was reconstructed as a triangular mesh using the marching cubes algorithm²⁰⁵, implemented via the scikit-image Python library.²⁰⁶ Representing the surface in this way accommodates the natural curvature and irregularity of the placenta and reduces the risk of overestimation and artefacts that can arise from voxel-based calculations.

6.2.2.2 Ratios of surface area to volume

Two complementary metrics — the surface-to-volume ratio (SVR) and the normalised, dimensionless UPI index — were derived from the reconstructed uteroplacental interface area A_{UPI} (cm²) and placental volume V (cm³).

The SVR was calculated as shown in Equation 4 (page 154), which provides a direct, physiologically interpretable measure of exchange surface per unit tissue volume.

However, it is more mathematically robust to use a dimensionless index because measures that retain physical units (like SVR) change simply with overall placental size (which will increase with gestational age) so differences between subjects may reflect scale rather than true differences in surface geometry. The formula in Equation 5 (page 154) was used to calculate the UPI index, which is dimensionless and therefore more mathematically robust, allowing comparison of different sized placentae in a fair and interpretable way.

6.2.2.3 Skeletonisation of the vascular network

To evaluate the uteroplacental vasculature, the vascular network was first skeletonised. After segmenting the vessels to obtain a binary mask (where voxels with any power Doppler signal are coded as 1 and those without coded as 0), we applied an axis-based thinning method to extract the vessel skeleton.²⁰⁷ This procedure generates a single, connected line of voxels that follows the geometric centreline of each vessel, representing points that are maximally distant from the vessel walls. Branch points are then identified as voxels connected to more than two neighbouring voxels, while voxels connected to only one neighbour correspond to vessel endpoints.

6.2.2.4 Identification of functional spiral artery vessel segments

In this study, we define functional spiral arteries (FSA) as vessel segments that arise from the uterine arcuate-radial network (that is, they begin with a branch point on the myometrial side of the UPI), traverse the UPI, and extend into the placental side of the UPI where they terminate. This is thought to include both the vessel itself, and the spiral artery jet as it enters the IVS. The jet component was defined as the portion of the vessel that extends between the UPI and the terminal point in the IVS. Segments were retained only if they overlapped with vessel regions and satisfied a minimum length threshold (≥ 1 mm), ensuring that the analysis focused on physiologically relevant branches and artefacts were removed.

6.2.2.5 Identification of functional arcuate and radial artery vessel segments

An algorithmic process was employed to categorise the remaining vessel segments (regions of a vessel between two branch points) as arcuate or radial arteries. Vessels were categorised as radial arteries if

they ran perpendicular to the UPI (within 45°) and had an unbroken line of connection to at least one FSA. The remaining unclassified vessels were identified as arcuate vessels if they were connected to a radial artery and had an unbroken line of connection to at least one FSA. It should be emphasised that these labels are operational rather than anatomical. Each category is assigned by geometric and topological rules applied to the skeletonised power Doppler signal, namely the orientation of a segment relative to the uteroplacental interface and its connectivity to other segments. No independent reference standard was available against which to confirm that a segment so labelled corresponds to the named vessel, and no such validation was attempted. The terms functional spiral artery, radial artery and arcuate artery should therefore be read throughout this chapter as denoting putative vessel classes derived from the imaging data, rather than as anatomically verified identifications.

6.2.2.6 *Evaluation of vessel morphology*

For each vessel segment, we derived the following metrics:

1. **Length:** determined by tracing the centreline and converting the number of voxels along this path into physical units using the voxel dimensions.
2. **Width:** obtained from the Euclidean distance transform of the segmented vessel volume, providing an estimate of the local vessel diameter at each point on the skeleton.
3. **Count:** the total number of distinct segments identified within the vessel group.

6.2.2.7 *Evaluation of functional spiral artery jet dispersion index*

For each FSA jet segment, the jet dispersion index (DI) was calculated. DI is a measure of how widely a vessel segment spreads relative to its length and it is calculated as shown in **Equation 6**, where V_{seg} is the segment volume and L_{seg} is the segment length of the centreline, both measured in voxels. This normalises the physical extent of a vessel segment (volume) by its longitudinal dimension (centreline length). A high dispersion index means the vessel segment is broader and more diffusely distributed, whereas a low value corresponds to a slender, narrow segment (**Figure 23**, page 154).

Equation 4: SVR

$$SVR (cm^{-1}) = \frac{A_{UPI}}{V}$$

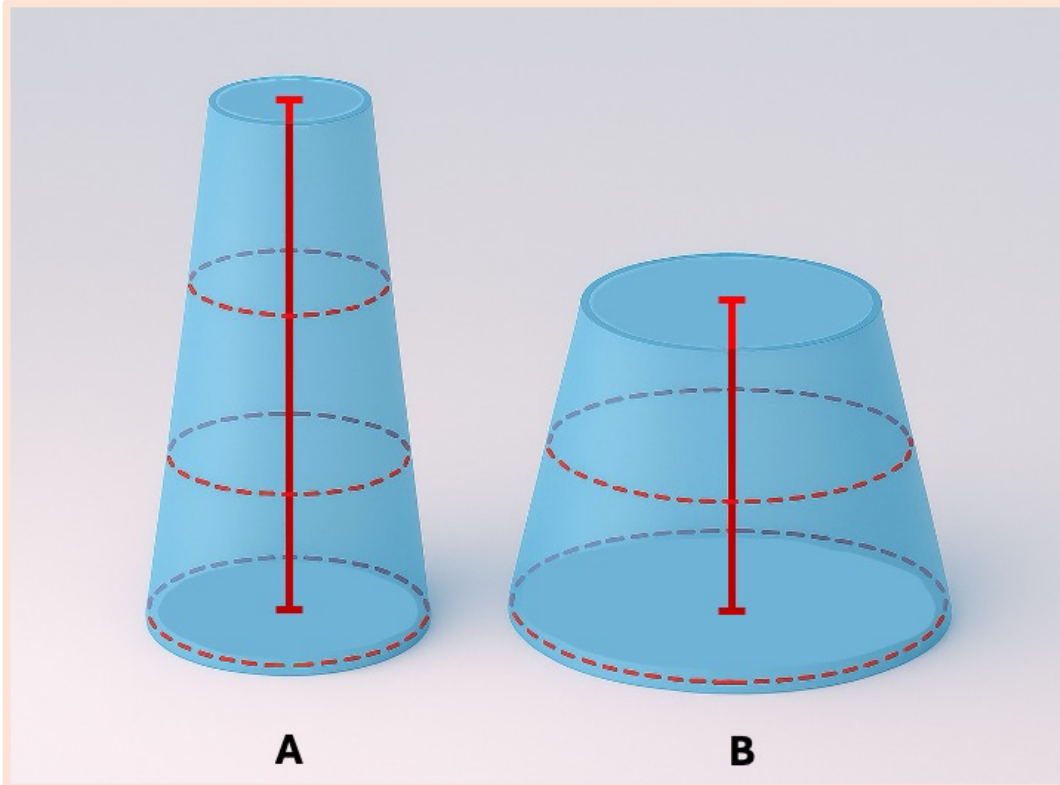
Equation 5: UPI index

$$I_{UPI} = \frac{A_{UPI}}{V^{2/3}}$$

Equation 6: Dispersion index

$$DI = \frac{V_{seg}}{L_{seg}}$$

Figure 23: Schematic representation of spiral artery jet morphology resulting in A) a low dispersion index and B) a high dispersion index. The solid red line represents the centreline of the vessel jet.



6.3 Statistical methods

Cohort characteristics and outcomes were described using simple descriptive statistics. For each novel marker, faceted boxplots were constructed to explore the changing distribution of values with advancing gestational age. For those that did not demonstrate any trend with advancing gestational age, simple histograms were produced to evaluate the distribution of data.

6.3.1 Handling of non-normal distributions

For variables that showed a trend of shifting distribution with advancing gestational age, the raw values were converted to z-scores using a simplified version of the LMS methodology described in section 3.2.4 (page 65) where only the default GAMLSS families were explored (including Box–Cox Cole and Green original, Box–Cox Power Exponential original, and Box–Cox t original distributions).^{158,162}

The remaining variables were normalised with $\log_{10}$ transformation if required and evaluated with simple histograms once again to visualise the results of the transformation.

6.3.2 Associations between novel parameters and adverse outcomes

Univariable binomial logistic regression was used to calculate the odds ratio for FGR and PE associated with an incremental increase of one for each marker. The Wald method was used to derive the 95% confidence intervals, and forest plots were constructed for visualisation of results. Thirteen candidate markers were each examined against two outcomes, giving twenty-six univariable comparisons. At a significance threshold of 0.05, rather more than one nominally significant association would be expected by chance alone. The analyses in this chapter are accordingly exploratory and hypothesis-generating, and the associations reported should be regarded as candidates for confirmation in an independent dataset rather than as established findings.

6.3.3 Subgroup analysis

Subgroup comparisons between successful and failed image processing were performed using the Wilcoxon Rank Sum test for continuous variables, and the Chi Square test (or Fisher's exact test if fewer than five events per group) for categorical variables. Statistical significance was set at $P < 0.05$.

Five subgroups were evaluated including those where processing failed for A) UPI metrics, B) arcuate artery metrics, C) radial artery metrics, D) FSA jet metrics (excluding failures of DI processing) and E) FSA jet dispersion index.

6.4 Results

A summary of key characteristics and outcomes for the combined cohort is shown in **Table 26**. Given that methodologies for extracting novel markers are still under refinement, processing failures occurred in certain cases. **Table 27** (page 158) provides a summary of data missingness for each variable.

Table 26: Summary of key characteristics and outcomes for the combined OxPLUS and FirstPLUS cohort (n = 6246)

Characteristic	Median (IQR) or N (%)
Age (years)	33.12 (29.88–35.99)
Height (cm)	166 (161–170)
Weight (kg)	68.00 (60.00–78.75)
BMI	24.93 (22.00–28.63)
Nulliparity	3131 (50.13)
Smoker at booking	186 (2.98)
Ethnicity	
White	4853 (77.70)
Black	535 (8.57)
South Asian	477 (7.64)
East Asian	114 (1.83)
Other or mixed	267 (4.27)
Gestational age at birth (days)	277 (272–284)
Pre-eclampsia	230 (3.68)
Gestational hypertension	269 (4.31)
Small for gestational age	642 (10.28)
Fetal growth restriction	362 (5.80)

Table 27: Summary of missing data for each novel marker (n = 6246)

Marker	Missing N (%)
UPI surface area (cm ²)	97 (1.55)
Surface-to-volume ratio	97 (1.55)
UPI index	97 (1.55)
Functional arcuate vessels	
Count	3686 (59.01)
Length	3686 (59.01)
Width	3686 (59.01)
Functional radial vessels	
Count	3675 (58.84)
Length	3675 (58.84)
Width	3675 (58.84)
Functional spiral artery jets	
Count	691 (11.06)
Length	691 (11.06)
Width	691 (11.06)
Dispersion index	3946 (63.18)

6.4.1 Exploration and transformation of data

Visual inspection of the faceted boxplots revealed distinct trends with advancing gestational age for some variables. UPI surface area demonstrated a progressive increase, whereas SVR showed a corresponding decline. In addition, the number of detectable functional spiral arteries (FSA) tended to rise with gestational age. These three metrics are illustrated in **Figure 24** (page 160). In all three cases, LMS analysis indicated the Box Cox Power Exponential original (BCPEo) distribution family provided the best fit and this method was therefore used to derive z-scores.

The faceted boxplots for variables that did not exhibit any apparent association with gestational age are presented in **Appendix 8** (page 222).

For those variables without gestational age–related trends, simple histograms are shown in **Figure 25** (page 161). The distributions of UPI index, arcuate artery mean length, radial artery mean length, arcuate artery mean width, radial artery mean width, FSA jet mean length, and FSA jet mean width were approximately normal, although in some cases the X-axis was extended by a small number of

extreme outliers. In contrast, arcuate artery count, radial artery count, and FSA jet mean dispersion index displayed varying degrees of positive skew and were therefore $\log_{10}$ transformed. The results of that transformation are presented as histograms in **Figure 26** (page 162), demonstrating that the transformation procedure was successful in approximating a normal distribution.

The fit of the LMS models for UPI surface area, SVR, and FSA jet count is presented as centile charts in **Appendix 9** (page 225). The distribution of the resulting z-scores is presented as histograms in **Figure 27** (page 163), demonstrating excellent normalisation for UPI surface area and SVR, but only moderately successful normalisation for FSA jet count where a cluster of extreme outliers is noted and the remaining data only approximate a normal distribution.

Figure 24: Faceted boxplots for novel placental markers plotted against gestational age for A) UPI surface area, B) surface volume ratio, and C) functional spiral artery jet count

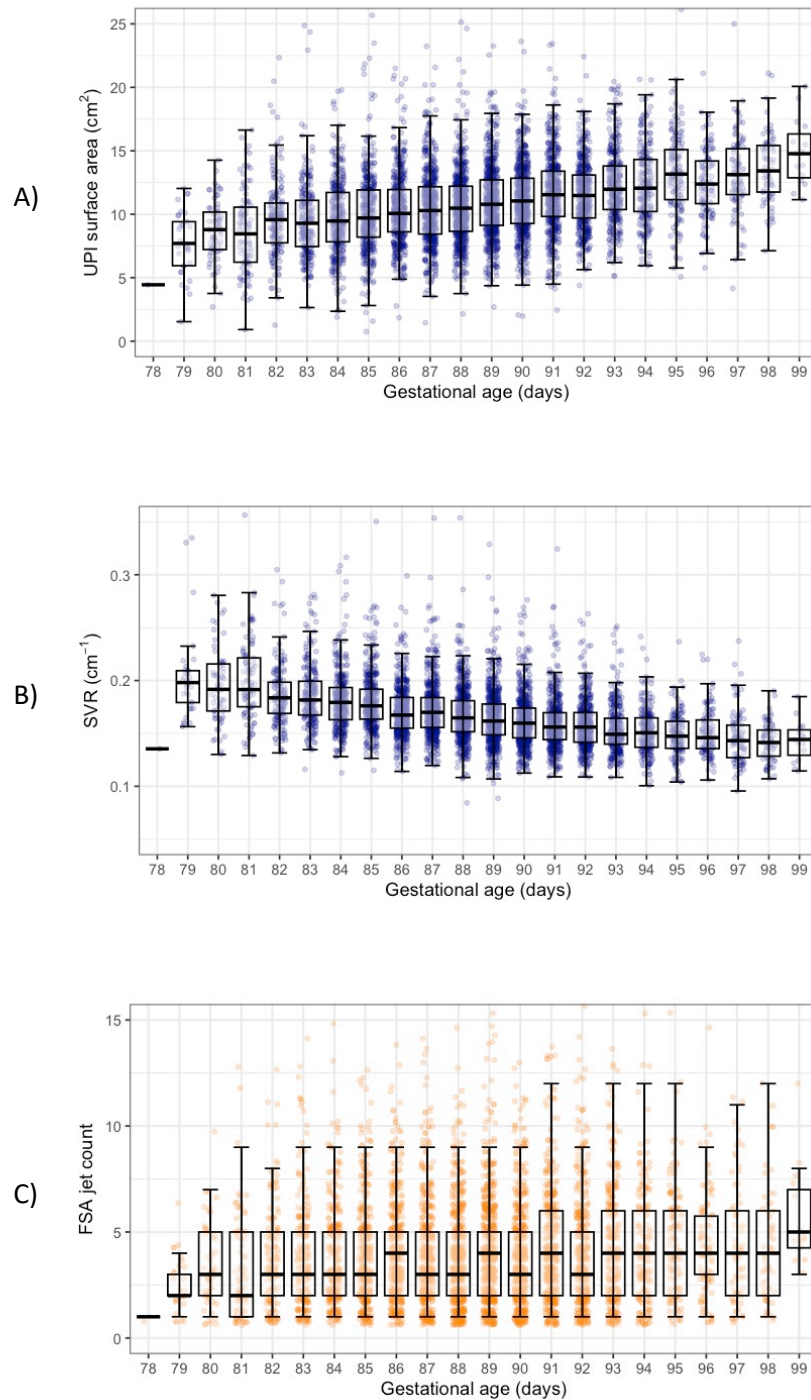


Figure 25: Histograms evaluating the distribution of novel placental markers

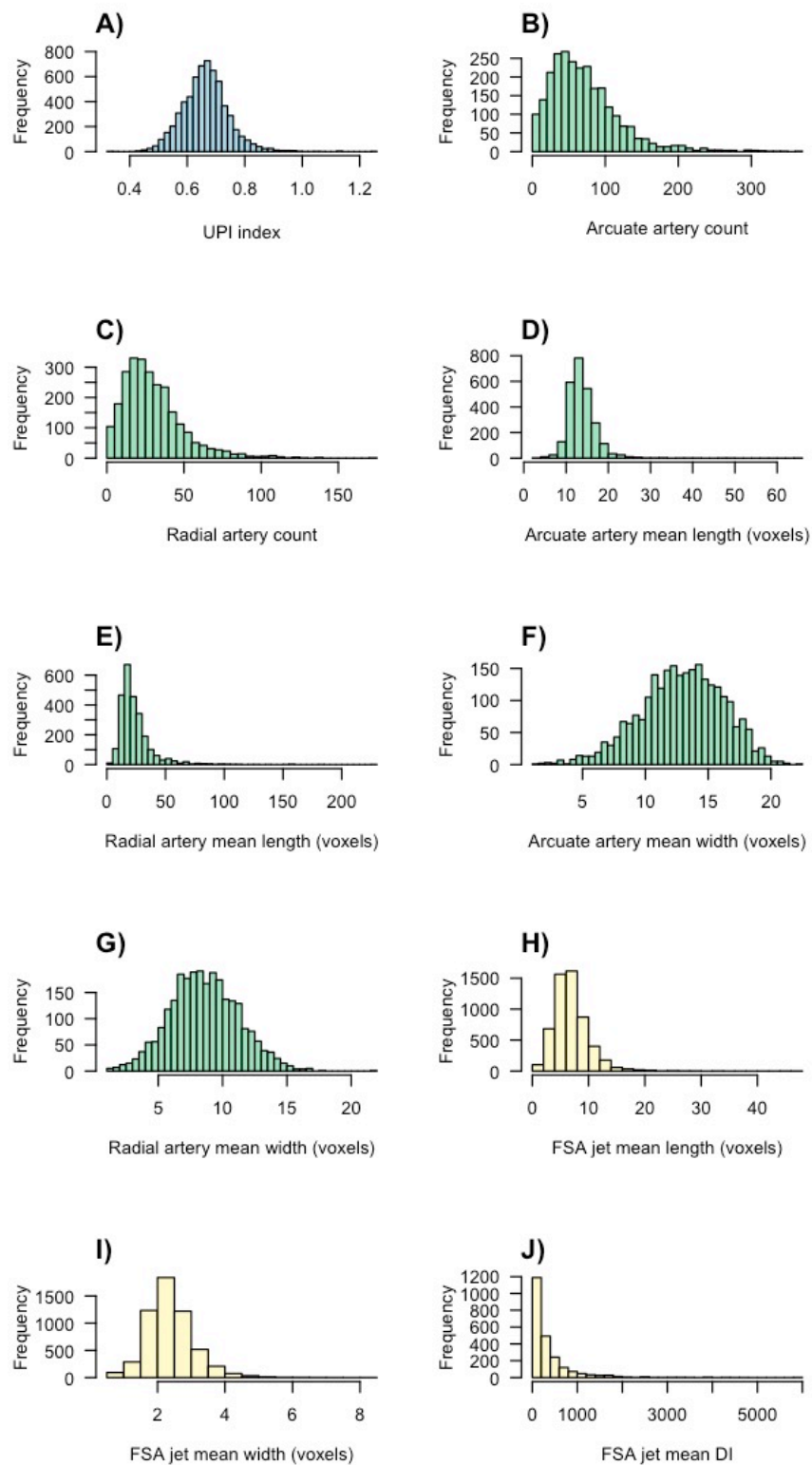


Figure 26: Histograms demonstrating the results of $\log_{10}$ transformation on distribution of novel markers for A) arcuate artery count, B) radial artery count, and C) FSA jet mean dispersion index

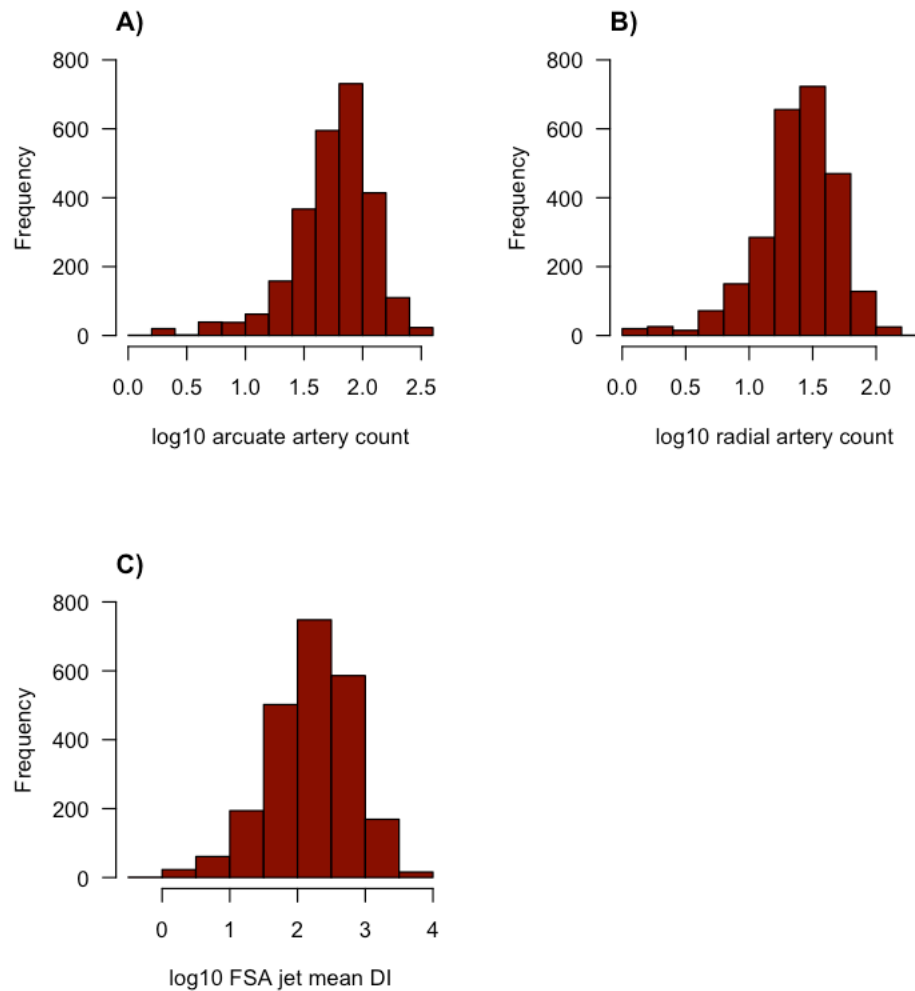
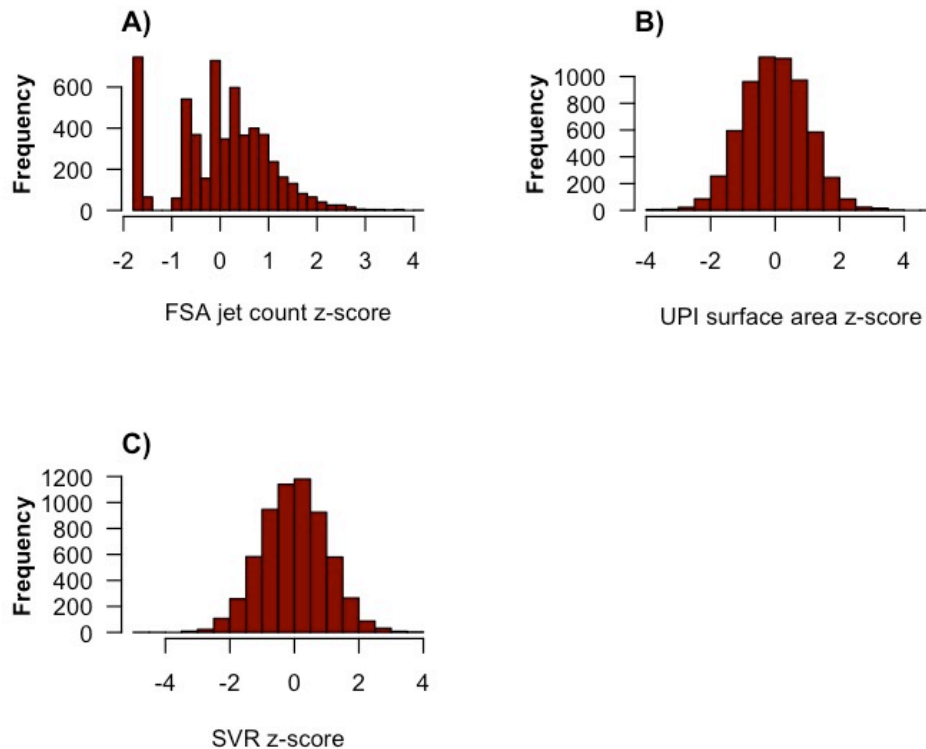


Figure 27: Histograms to evaluate the distribution of z-scores derived from LMS modelling for A) functional spiral artery jet count, B) uteroplacental interface surface area, and C) surface volume ratio



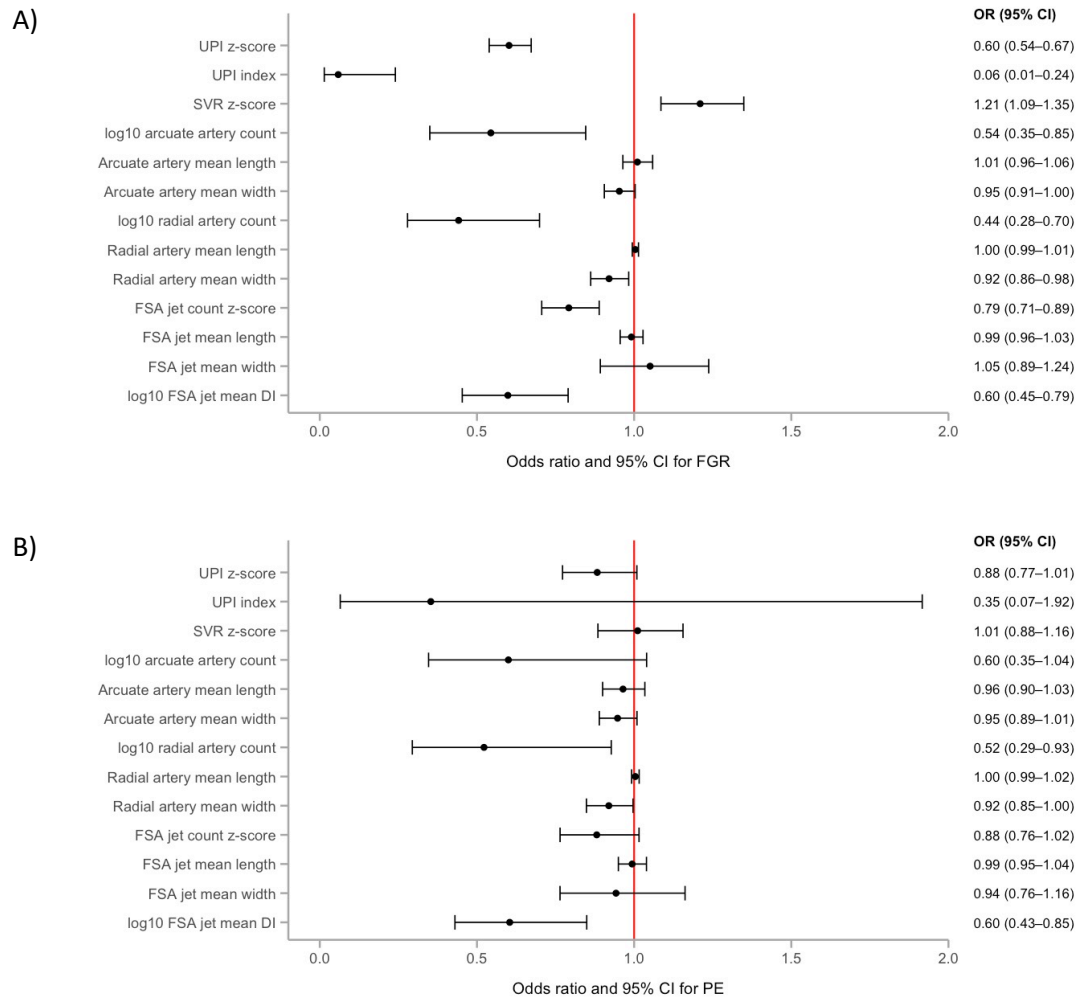
6.4.2 Associations between novel markers and adverse outcomes

The forest plots showing odds ratios and 95% confidence intervals for each adverse outcome associated with each novel marker are presented in **Figure 28** (page 164).

For univariable prediction, statistically significant associations with FGR were observed for the UPI z-score, UPI index, SVR z-score, $\log_{10}$ arcuate artery count, $\log_{10}$ radial artery count, radial artery mean width, FSA count z-score, and $\log_{10}$ FSA jet mean DI.

For univariable prediction of PE, only $\log_{10}$ radial artery count and $\log_{10}$ FSA jet mean DI were statistically significant. However, a borderline significant result was also observed for radial artery mean width (OR 0.92, 95% CI 0.85–1.00).

Figure 28: Forest plots showing odds ratios for A) fetal growth restriction and B) pre-eclampsia associated with an incremental increase of one for each novel marker



6.4.3 Subgroup analysis

The complete subgroup analysis is available in **Appendix 10** (page 226). No statistically significant differences were observed for age, height or smoking in any of the missing data subgroups compared with cases where processing was successful. Nor were any differences observed in the outcomes of PE or PIH.

However, failed processing of UPI metrics was associated with an increased probability of SGA (10.19% in the success group versus 17.53% in the failure group, $P = 0.029$) and FGR (5.65% in the success group versus 16.49% in the failures group, $P < 0.001$). Failed processing of arcuate artery metrics was associated with slightly greater weight (median of 67 kg in the success group versus 69 kg in the failure group, $P < 0.001$), greater BMI (24.39 versus 25.00, $P < 0.001$), nulliparity (45.66% versus 53.23%, $P < 0.001$), and ethnicity (particularly with respect to black ethnicity, 6.95% in the success group versus 9.69% in the failure group, $P < 0.001$). Virtually identical results were observed for failed processing of radial artery metrics, although a statistically significant difference in gestational age was observed despite the median being 277 days in both groups (IQR 273–283 versus 272–284, $P = 0.017$) which while statistically significant, does not imply clinical significance. Failed processing of the FSA jet related metrics (excluding dispersion index) also revealed very similar discrepancies in weight and BMI, although the ethnic distribution did not differ ($P = 0.619$). Failed processing of FSA jet dispersion index was also associated with a similar difference in weight and BMI. Black ethnicity was again overrepresented in the failure group (6.70% in the success group compared with 9.66% in the failure group), and in this case the difference was statistically significant ($P < 0.001$).

6.5 Discussion and conclusion

This work has shown that it is possible, using OxNNet, to derive hitherto unexplored candidate markers of FGR and PE from the properties of the uteroplacental interface and the uteroplacental vasculature. All the UPI-derived markers demonstrated statistically significant associations with FGR but not PE. This may reflect differing pathological processes between the two conditions but could also be in part due to the lower event rate of PE compared with FGR contributing to lack of statistical significance, and UPI z-score was very close to statistical significance with respect to its relationship with PE (OR 0.88, 95% CI 0.77–1.01). The arcuate and radial artery counts were both associated with FGR, but only the radial artery count had a statistically significant association with PE. Radial artery width was also associated with FGR. The corresponding estimate for PE was of similar magnitude but did not reach statistical significance, the upper bound of the confidence interval being one (OR 0.92, 95% CI 0.85–1.00).

The FSA jet count and mean dispersion index were associated with FGR and dispersion index also had a statistically significant association with PE.

6.5.1 Findings in the context of current understanding

6.5.1.1 Uteroplacental interface

The findings in relation to UPI-derived metrics complement what is already known. In a 2016 prospective cohort study, Salavati *et al.* examined the placenta in 2120 cases after delivery, and found an inverse correlation between placental surface area and uterine artery PI z-score at 20 weeks' gestation (a known important predictor of both FGR and PE), and also that there was an association between the lowest decile of placental surface area and FGR.²⁰⁸ Similarly, in a case control study comparing 70 FGR placentae with 70 healthy controls, Liu *et al.* (2021) measured the surface area of the maternal side of the placenta shortly after birth, and found that this was significantly smaller in the FGR cases.²⁰² Suri *et al.* (2018) undertook a cohort study in which 14 cases of pre-eclampsia were compared with 301 healthy controls. The dimensions of the basal plate of the placenta were evaluated using two-dimensional B-mode ultrasound late in the first trimester, and cases that developed PE were found to have a smaller basal plate surface area ($P < 0.01$).²⁰⁹ There is also lab-based stereological evidence to suggest that the exchange surface area itself (i.e. the total villous surface area) is reduced

in association with FGR, but this was not observed for PE.^{210,211} In comparison to existing research, this study provides a much larger and relatively unselected dataset, and focuses on first trimester UPI metrics — this is important because this is the gestation at which screening for PE and FGR must take place in order for prophylactic strategies to have maximum effect.

6.5.1.2 Arcuate and radial arteries

Ong et al. (2005) investigated whether the structure of myometrial radial arteries differs between normal pregnancies and those complicated by PE or FGR.²¹² Using pressure myography, they demonstrated that vessels from PE pregnancies exhibited a smaller lumen, increased wall thickness, and a higher wall-to-lumen ratio compared with arteries from uncomplicated pregnancies. In contrast though, arteries obtained from FGR pregnancies showed no convincing structural differences relative to controls.²¹² This is contrary to work presented in this chapter, where an association was observed with both outcomes (although the upper bound of the 95% confidence interval was one, with respect to PE). Any such comparison should, however, be treated with caution, since it presumes that the vessel segments labelled as radial in the present analysis correspond to the myometrial radial arteries studied by pressure myography. That correspondence is assumed rather than demonstrated, and the same caveat applies to the comparisons with arcuate and spiral artery data that follow.

Clark et al. (2018) developed a comprehensive computational model of the uteroplacental vasculature to explore the origins of abnormal uterine artery Doppler findings.⁶⁹ By integrating the myometrial arterio-venous anastomoses into their model, together with detailed structural representations of each vascular component, the authors were able to apply wave transmission theory to predict UTA Doppler waveforms across the whole vascular system. Their results indicate that high resistance uterine artery Doppler indices are more consistent with abnormal remodelling throughout the uterine vasculature rather than isolated disruption at the level of the spiral arteries. Indeed, incomplete spiral artery remodelling alone appeared insufficient to generate abnormal waveforms, as increased resistance could be offset by upstream anastomotic pathways. Importantly, the findings suggest that the radial arteries may exert a greater influence on uteroplacental flow dynamics than previously appreciated.⁶⁹ This is consistent with the findings of this chapter, although stronger associations with arcuate and radial artery count were observed with FGR compared with PE.

Allerkamp et al. (2020) provide rare quantitative data on the physiological evolution of uterine vascular remodelling and trophoblast plugging during the first half of pregnancy, drawing on digitised serial sections from historical in situ specimens.²¹³ Across 6–20 weeks’ gestation, they observed substantial dilation of arcuate, radial and spiral arteries (approximately 2.6 to 4.3 fold) indicating that adaptation of the uterine vascular tree is both progressive and not complete by the end of the first trimester. This supports the notion that disruption of these observed changes to the arcuate and radial vessels might have relevance to disease processes.

6.5.1.3 *Spiral arteries*

Ultrasound assessments of individual spiral artery inflow jets report elevated resistance patterns in small-for-gestational-age pregnancies^{61,62}, which supports the well-recognised link between abnormal spiral artery remodelling and the development of FGR and PE. This is supported by extensive histopathological evidence^{41,42} and more recently by computational studies illustrating the resulting disturbances in blood flow.^{214,215} Narrow, insufficiently remodelled spiral arteries are thought to generate high-velocity streams of blood as they enter the intervillous space, potentially damaging the villous surface — the so-called ‘hose-effect’.⁵³ This is consistent with the observation from this chapter that the FSA jet mean dispersion index is associated with both FGR and PE, with higher dispersion index associated with lower risk of the adverse outcomes which is the expected direction of effect.

6.5.2 Strengths and limitations

A key strength of this study lies in its comparatively large cohort, drawn from two independent centres, alongside the application of machine-learning methods to automate image segmentation and analysis. These features enhance both the reproducibility and scalability of the approach.

It is important, however, to acknowledge several limitations, particularly given that the markers introduced here are novel, exploratory, and still undergoing refinement. The most significant constraint is the high proportion of images that could not be successfully processed. While a subset of these failures may represent a genuine lack of measurable inflow, many are likely attributable to technical factors, including diminished image quality caused by signal attenuation in participants with higher BMI. The finding that nulliparity and Black ethnicity were associated with unsuccessful

processing is noteworthy, especially as both characteristics are recognised risk factors for PE and FGR.^{139,216-219}

The scale of processing failure reduces the volume of evaluable data and heightens the potential for bias. This underscores the need for further work not only to enhance the robustness of the image-processing pipeline — and potentially optimise image acquisition to improve success rates — but also to investigate more thoroughly the factors underpinning differential processing success across ethnic groups.

An additional limitation is that exact anatomical validation of these novel measurements remains unfeasible currently. Although the intended targets are the named vessels of the placental bed, minor segmentation deviations may result in the inclusion of basal arteries or other vessels adjacent to the uteroplacental interface, rather than exclusively the spiral arteries. Furthermore, while the distinction between arcuate and radial vessels is based on the established anatomy of the myometrial vascular plexus, it is a pragmatic approach, and the extent of possible vessel misclassification cannot currently be determined.

Finally, the absence of correction for aspirin use within the study cohort may lead to underestimation of the strength of association between predictors and outcomes. Low-dose aspirin is a well-established preventive strategy for PE — particularly preterm disease — and there is also some evidence supporting its role in reducing the risk of FGR (see paragraph **1.3.2**, page 30).^{83,220} Among those taking aspirin, some cases of PE and FGR will have been prevented. Consequently, the observed associations between the imaging markers and adverse outcomes may be attenuated, which could partly explain why some relationships did not reach statistical significance. At this early, proof-of-concept stage, this limitation can simply be noted, but future studies evaluating these markers should incorporate adjustment for aspirin exposure.

6.5.3 Conclusion

This work demonstrates the feasibility and potential clinical relevance of deriving novel, quantitative descriptors of uteroplacental architecture and vascularity from first-trimester ultrasound using automated machine-learning methods. Although these markers remain exploratory and require further refinement, the consistent associations observed with FGR — and, for selected measures, with PE — suggest that some of these markers, captured early in gestation, may hold genuine predictive value.

Importantly, these findings are consistent with current understanding of uteroplacental pathophysiology, and raise the possibility that vessels of the arcuate and radial classes contribute to adverse pregnancy outcomes. At the same time, the limitations identified — particularly the high rate of image-processing failure, uncertainty in precise anatomical attribution, and lack of adjustment for aspirin exposure — indicate that further development is required before these markers can be translated into clinical practice. Nevertheless, this proof-of-concept work provides a strong foundation for future research aimed at improving early screening and pathophysiological insight for FGR and PE.

The extent of missing data reinforces this exploratory framing and bears directly on how the associations should be read. Image processing was unsuccessful in 59% of cases for the arcuate metrics, 59% for the radial metrics and 63% for the jet dispersion index, so the associations reported for these markers rest on a minority of the cohort. That minority is unlikely to have been selected independently of outcome risk, since unsuccessful processing was itself associated with higher maternal body mass index, nulliparity and Black ethnicity, each of which is associated with fetal growth restriction and pre-eclampsia. The direction of any resulting bias cannot be determined from these data, and the associations should therefore be regarded as candidates for confirmation in a dataset with more complete processing rather than as estimates of effect.

[7] REPRODUCIBILITY OF OXNNET METRICS

7.1 Introduction

The reproducibility of pregnancy ultrasound measurements depends on modality, operator experience, and anatomical target, but ultrasound is generally considered especially prone to inter-operator variability among medical imaging methods. Examples relevant to this thesis include first trimester nuchal translucency measurement and uterine artery Doppler velocimetry, both of which have applications for screening in early pregnancy.

Nuchal translucency (NT) measurements, for example, demonstrate high reproducibility when conducted by well-trained sonographers, with intra-operator intraclass correlation coefficients (ICC) generally exceeding 0.83 and inter-operator ICCs among experienced practitioners ranging from 0.74 to 0.95. However, inter-operator ICCs can be as low as 0.51 among less experienced operators.²²¹ Additionally, equipment calibration plays a critical role in NT measurement accuracy, and variability between machine-probe combinations may exert an even greater influence on reproducibility than operator expertise.²²² First trimester uterine artery Doppler measurements generally demonstrate intra- and interobserver ICCs in the poor to moderate range, with reliability further reduced in women with higher BMI.²²³⁻²²⁵

OxNNet metrics, especially placental volume and svFMBV, have demonstrated potential as first trimester screening tools for FGR and PE. To support the potential clinical implementation of OxNNet measurements, it is essential to thoroughly evaluate their limitations; therefore, an assessment of their reproducibility is warranted.

7.1.1 Aims

To evaluate the intra- and inter-operator reproducibility of OxNNet-derived placental metrics and of the novel markers derived in Chapter 6.

7.1.2 Statement of personal contribution

To complete this work, I have contributed to the data collection (performing placental ultrasound scans) alongside Dr. Theophilus Adu-Bredu, research sonographer. Image processing work was completed by Dr. Mohammadreza Soltaninejad. I have completed all subsequent statistical analyses myself.

7.2 Methodology

7.2.1 Data collection protocol

The first-trimester placental ultrasound protocol was repeated in participants opportunistically selected from the OxPLUS study cohort. For assessment of intra-operator reproducibility, each participant was scanned twice by the same sonographer. For evaluation of inter-operator reproducibility, the scan was repeated by a second operator.

A new study record was generated on the ultrasound machine for every examination, and the system was reset to its default presets prior to each scan. This included re-establishing the standard greyscale and power Doppler gain settings. During inter-operator assessments, the second operator was blinded to the machine settings applied by the first operator to minimise bias.

7.2.2 Sample size determination

The required sample size was determined using the `prec_icc` function from the `presize` package in R, following the approach outlined by Bonett (2002).²²⁶ Assuming an expected intraclass correlation coefficient (ICC) of 0.8 and a confidence level of 95%, a sample of 28 cases was estimated to be sufficient to achieve a confidence interval width of 0.3. Larger sample sizes would allow for the estimation of lower ICC values with correspondingly narrower confidence intervals.

7.2.3 Selection of variables for analysis

OxNNet metrics were considered for reproducibility based on whether a statistically significant association has been demonstrated with FGR or PE. Placental volume and svFMBV were therefore included, with svFMBV5 chosen specifically due to its stronger reported association with FGR and PE compared with other svFMBV parameters. In addition, novel markers identified in chapter 6 as having statistically significant associations with FGR or PE were incorporated, namely the UPI surface area, UPI index, and SVR.

Arcuate artery count, radial artery count, radial artery width, and FSA jet count were excluded from further evaluation. These parameters were prone to processing failure (as discussed in paragraph 6.5.2, page 168), resulting in insufficient numbers of matched data pairs (fewer than three in all cases) to permit meaningful reproducibility assessment.

7.2.4 Statistical approach

ICC values and 95% confidence intervals were estimated using the `icc` function from the `irr` package in R, which is based on the original method by Shrout and Fleiss (1979) and relies on an assumption that the data are normally distributed.²²⁷ To ensure normality of distribution, placental volume and `svFMBV5` values were converted to z-scores as described in chapter 3 (page 62), and UPI area and SVR were converted to z-scores using the method described in chapter 6 (page 147). The UPI index naturally has a normal distribution as demonstrated in **Figure 25** (page 161), and so no transformation was necessary.

Because the same operators performed the repeat examinations, subjects and operators were fully crossed and a two-way model was therefore appropriate. The coefficient was estimated as a measure of absolute agreement rather than relative consistency, based on single measurements per subject. For inter-operator reproducibility this corresponds to a two-way random effects model with absolute agreement and single measures, denoted ICC(2,1) in the notation of Shrout and Fleiss²²⁷ and ICC(A,1) in that of McGraw and Wong²²⁸; the operators are treated as random because they represent a sample of those who might perform the examination in practice. For intra-operator reproducibility the two occasions are the only occasions of interest, so a two-way mixed effects model applies, again with absolute agreement and single measures, ICC(A,1). The absolute-agreement estimate is identical whether the second factor is treated as random or fixed, the distinction affecting generalisation rather than computation. Estimates were obtained with the `model`, `type` and `unit` arguments of the `icc` function set to `twoway`, `agreement` and `single` respectively. Confidence intervals at the 95% level were derived using the F distribution-. For each variable of interest, ICC values were computed along with their corresponding confidence intervals, providing an index of reliability across repeated measurements.

Bland–Altman plots were constructed to evaluate agreement between repeated measurements. For each variable, the mean of the paired values was plotted against their difference. The mean difference (bias) was calculated, together with the standard deviation of the differences. Limits of agreement (LOA) were defined as the mean difference ± 1.96 times the standard deviation, representing the range within which 95% of differences would be expected to lie. The proportion of observations falling within, above, and below these limits was also reported.

7.2.4.1 Interpretation of intraclass correlation coefficients

There is no universally accepted standard for what qualifies as an acceptable ICC. What is considered acceptable depends on the context in which the measurement is used, and it is important to consider the range of the 95% confidence interval. Generally, ICC values above 0.5, 0.75, and 0.9 are interpreted as indicating moderate, good, and excellent reliability, respectively.²²⁹

7.3 Results

There were 52 cases available for intra-operator reproducibility analysis, and 47 cases for inter-operator analysis. The demographic characteristics of each group are shown in **Table 28** (page 176).

Table 28: Summary of demographic characteristics for intra-operator and inter-operator reproducibility analysis cohorts

Characteristic	Intra-operator analysis (n = 52) Median (IQR) or N (%)	Inter-operator analysis (n = 47) Median (IQR) or N (%)
Age (years)	33.68 (30.55–36.45)	33.18 (30.22–36.05)
Height (cm)	166 (160.75–169.25)	167 (161.5–169)
Weight (kg)	65.5 (58.75–78.25)	68 (60.5–79)
BMI	24.5 (22–29.25)	26 (22–29.5)
Nulliparity	26 (50)	26 (55.3)
Smoker at booking	3 (5.8)	3 (6.4)
Ethnicity		
White	43 (82.7)	40 (85.1)
Black	1 (1.9)	1 (2.1)
South Asian	7 (13.5)	5 (10.6)
East Asian	1 (1.9)	1 (2.1)
Other or mixed	0 (0)	0 (0)

The intra-operator agreement analysis (**Table 29**, page 178) demonstrated generally high reproducibility, with ICCs ranging from 0.57 to 0.85. The strongest agreement was observed for the placental volume z-score (ICC 0.85, 95% CI 0.76–0.91) and UPI z-score (ICC 0.85, 95% CI 0.75–0.91), followed closely by the SVR z-score (ICC 0.80, 95% CI 0.67–0.88), while the UPI index showed moderate reliability (ICC 0.68, 95% CI 0.51–0.81) and the svFMBV5 z-score had the lowest reproducibility (ICC 0.57, 95% CI 0.36–0.73). Mean differences across parameters were close to zero, indicating minimal systematic bias, and the majority of measurements (92–94%) fell within the limits of agreement.

The inter-operator agreement analysis (**Table 30**, page 178) revealed slightly lower reproducibility, with ICCs ranging from 0.58 to 0.77. The SVR z-score (ICC 0.77, 95% CI 0.63–0.87), volume z-score (ICC 0.73, 95% CI 0.57–0.84) and UPI z-score (ICC 0.72, 95% CI 0.54–0.83) showed the most consistent agreement between operators, while the UPI index demonstrated moderate agreement (ICC 0.65, 95% CI 0.45–0.79), and svFMBV5 again had the weakest reproducibility although still within the range

considered acceptable (ICC 0.58, 95% CI 0.35–0.74). Mean differences were negligible across all parameters, suggesting no systematic operator bias, and the majority of values (94–98%) were within the limits of agreement.

The results of both intra- and inter-operator analyses are visually represented as Bland-Altman plots in **Figure 29** (page 179).

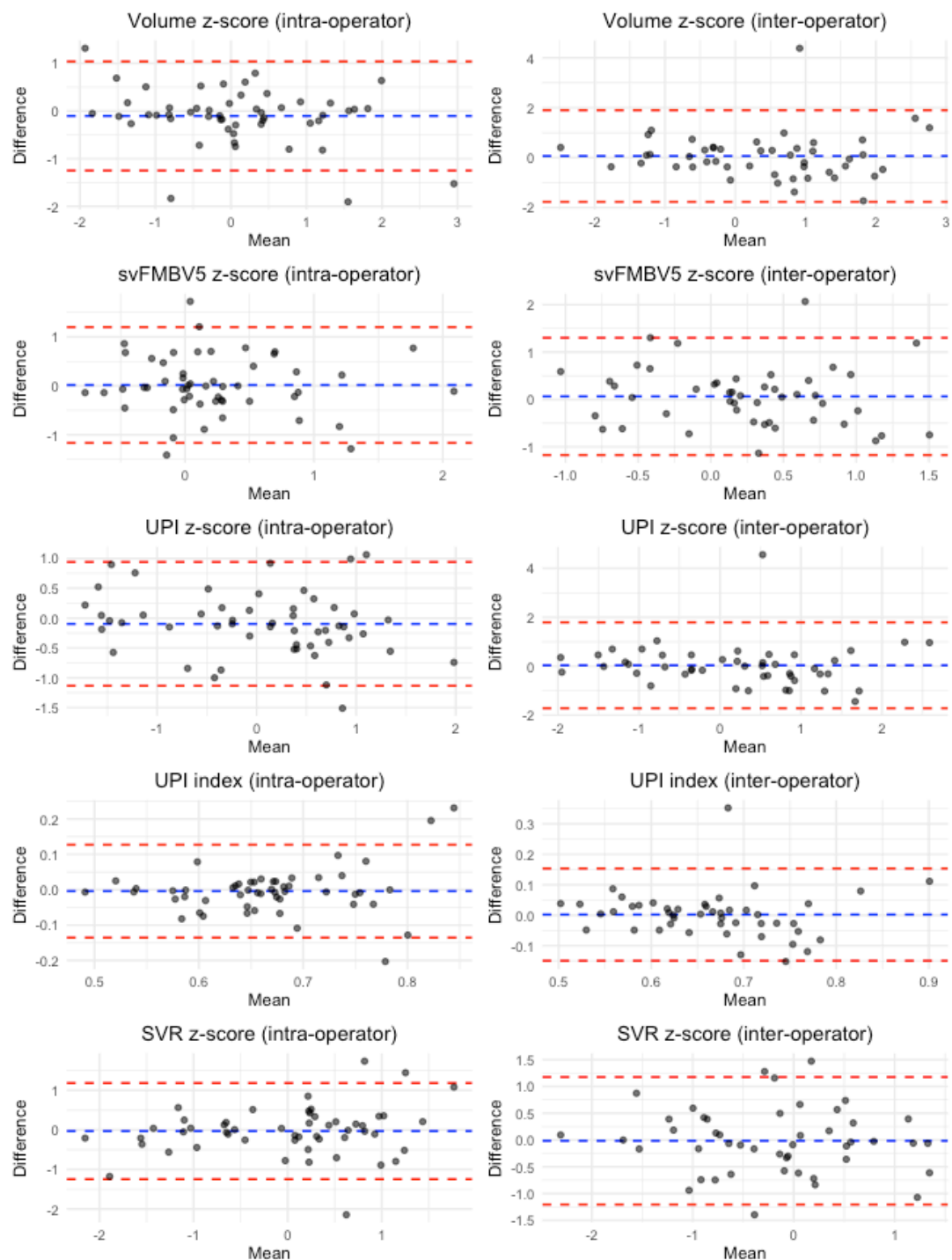
Table 29: Intra-operator agreement analysis

Parameter	ICC (95% CI)	Mean difference	Limits of agreement	Within limits (%)	Above limits (%)	Below limits (%)
Volume z-score	0.85 (0.76–0.91)	-0.1	-1.25–1.04	92.31	1.92	5.77
svFMBV5 z-score	0.57 (0.36–0.73)	0.02	-1.17–1.20	92.31	3.85	3.85
UPI z-score	0.85 (0.75–0.91)	-0.10	-1.13–0.94	94.23	3.85	1.92
UPI index	0.68 (0.51–0.81)	0	-0.14–0.13	94.23	3.85	1.92
SVR z-score	0.80 (0.67–0.88)	-0.03	-1.25–1.18	94.23	3.85	1.92

Table 30: Inter-operator agreement analysis

Parameter	ICC (95% CI)	Mean difference	Limits of agreement	Within limits (%)	Above limits (%)	Below limits (%)
Volume z-score	0.73 (0.57–0.84)	0.06	-1.77–1.90	97.87	2.13	0
svFMBV5 z-score	0.58 (0.35–0.74)	0.06	-1.17–1.30	95.74	4.26	0
UPI z-score	0.72 (0.54–0.83)	0.04	-1.71–1.79	97.87	2.13	0
UPI index	0.65 (0.45–0.79)	0	-0.15–0.15	95.74	2.13	2.13
SVR z-score	0.77 (0.63–0.87)	-0.01	-1.20–1.18	93.62	4.26	2.13

Figure 29: Bland-Altman plots for five OxNNet-derived first trimester placental markers demonstrating intra-operator reproducibility (left column) and inter-operator reproducibility (right column)



7.4 Discussion and conclusion

This study provides the first systematic evaluation of the reproducibility of OxNNet-derived placental metrics addressing both intra- and inter-operator agreement. Overall, the findings demonstrate that most OxNNet parameters achieve good intra-operator reproducibility and moderate inter-operator reproducibility. Intra-operator agreement was generally stronger than inter-operator agreement, which is consistent with expectations for ultrasound-based measurements where operator technique and probe positioning can introduce variability.^{221,223-225}

Inter-operator agreement was slightly lower than intra-operator agreement for most parameters. This reduction reflects the inherent variability of ultrasound imaging between operators, a well-recognised limitation in obstetric sonography.²²³⁻²²⁵ Nevertheless, reproducibility remained good for SVR, moderate for placental volume, the UPI z-score and the UPI index, and lowest for svFMBV5, with negligible mean differences across all parameters indicating minimal systematic bias.

Importantly, the reproducibility observed here compares favourably with established first-trimester ultrasound markers such as nuchal translucency and uterine artery Doppler velocimetry, which have historically demonstrated similar ranges of ICC values depending on operator experience and patient characteristics.^{221,223-225} This suggests that OxNNet-derived metrics are at least as reliable as current clinical standards, while offering novel insights into placental development.

7.4.1 Findings in the context of current understanding

As OxNNet parameters are so novel, there is no existing literature for comparison that is entirely apposite. However, comparison can be made against other techniques for placental segmentation and evaluation of placental vascularity.

Cabezas López et al. (2016) conducted a study involving 69 singleton pregnancies to evaluate parameters derived using VOCAL™.²³⁰ For placental volume, they reported an intra-observer ICC of 0.97 (95% CI 0.95–0.98) and an inter-observer ICC of 0.71 (95% CI 0.53–0.82). The vascularisation index (VI), flow index (FI), and vascularisation flow index (VFI) also demonstrated strong reproducibility, with intra-operator ICCs of 0.98 (95% CI 0.96–0.99), 0.93 (95% CI 0.89–0.96), and 0.97 (95% CI 0.96–0.98), respectively. Inter-operator ICCs were similarly high at 0.96 (95% CI 0.94–0.98) for VI, 0.89 (95% CI 0.82–0.93) for FI, and 0.95 (95% CI 0.91–0.97) for VFI. These findings suggest excellent reproducibility

of measurements obtained with VOCAL™, but critically these results reflect repeat analyses performed on a single image per participant. Consequently, the reproducibility demonstrated pertains to the segmentation tool itself rather than the variability inherent in first trimester placental ultrasound image acquisition. In comparison, the work presented in this chapter is a better reflection of reproducibility during real-world use because it assesses the variables of both image acquisition and image processing.

Similar findings were reported by Nardoza et al. (2009), who examined repeated measures of placental volume derived using VOCAL™ in a subgroup of 16 patients from a larger study. In this case, two operators performed analyses on a single ultrasound image per patient. While reproducibility was demonstrated, the reliance on a single image substantially limits the applicability of these results to real-world clinical practice, where variability in image acquisition is a critical factor.²³¹

By contrast, Huster et al. (2010) evaluated 30 pregnancies using true repeated measures and found that placental volume derived with VOCAL™ achieved an inter-observer ICC of 0.88 (95% CI 0.75–0.94).²³² Intra-observer reproducibility varied between operators, with ICCs of 0.95 (95% CI 0.89–0.98) and 0.84 (95% CI 0.68–0.92), respectively. Vascular indices (VI, FI, and VFI) also demonstrated good inter-operator agreement within a similar range. However, a notable methodological limitation was the use of fixed power Doppler gain settings across all acquisitions. Although this approach likely improved reproducibility — given that gain settings strongly influence Doppler signal detection¹¹⁴ — it fails to account for other beam path factors such as maternal body habitus and placental depth. Standardising gain settings therefore risks artificially inflating reproducibility while undermining the validity of vascularity assessments. The broader challenges associated with FI, VI, and VFI in this context have already been discussed in paragraph 1.4.2 (page 34).

None of these studies mention whether the data met the normality assumption for ICC calculation.

7.4.2 Strengths and limitations

This study is strengthened by its robust repeated-measures design, which included full ultrasound machine resetting between scans. This ensured that all aspects of image acquisition and processing were evaluated with respect to reproducibility, thereby providing a more accurate assessment of real-world performance. The blinding of the second assessor further mitigated the risk of inadvertent bias,

reducing the likelihood of underestimating discrepancies in measurements. Relative to comparable literature, the study is also reinforced by its relatively large sample size and its rigorous approach to data normalisation when required.

Further exploration of the reasons underlying the relatively weaker performance of svFMBV5 would be valuable in determining how this measure might be optimised, both in terms of reproducibility and predictive utility. A key limitation of the present work is the absence of recorded information on power Doppler gain settings for each scan, which might otherwise have been included in this reproducibility analysis. Variation in gain selection is likely to exert a significant influence on the final output of all svFMBV measures. Despite evidence that the subnoise gain selection method is highly reproducible¹³¹, variation in this measure between scans might account for some of the variability observed. Evaluating the intra- and inter-rater reproducibility of power Doppler gain settings within this cohort would therefore be particularly informative. At present, such data cannot be readily extracted from the kretzfile metadata, as this information is embedded within proprietary coding. Manual case-by-case extraction, however, remains feasible and represents a worthwhile avenue for future research. Establishing whether the subnoise gain selection method is genuinely as reproducible as claimed is a critical step towards clarifying the overall performance and utility of svFMBV measures.

A limitation of the agreement analysis should be made explicit. The proportion of measurements falling within the limits of agreement is not an independent criterion of acceptability, because those limits are derived from the same differences and will by construction contain approximately 95% of them, whatever the magnitude of the underlying variability. What matters clinically is the width of the limits, judged against the difference that would be tolerable in use, and no such threshold was pre-specified for these metrics.

Interpreted on that basis the limits are wide. For intra-operator agreement they span roughly 2.3 to 2.4 z-score units for placental volume, svFMBV5 and the surface-to-volume ratio, and for inter-operator agreement 3.7 units for placental volume and 3.5 for the UPI z-score — in each case exceeding twice the standard deviation of the reference distribution from which the z-scores are derived. A woman whose true placental volume lay on the 50th centile could therefore return a repeat measurement anywhere between approximately the 11th and the 85th centile if rescanned by the same operator, and between the 4th and the 97th centile if rescanned by a different one. Differences

of that size are not negligible in relation to the scale on which these metrics are interpreted, being of the same order as the z-score thresholds of ± 1 , ± 2 and ± 3 used for multi-rule quality control in Chapter 4.

Reproducibility of this order is adequate for the group-level analyses presented here, in which measurement error is averaged across large cohorts. Whether it would be sufficient for an individual measurement cannot be determined from these data alone, since that judgement depends on the difference considered tolerable for a given clinical purpose, and no such threshold has yet been defined for these metrics. Establishing one, and determining whether a single acquisition or an average of repeated acquisitions would be needed to meet it, is a necessary step before these measures are applied in individual care.

Finally, this study employed the same ultrasound machine and probe across all cases. This approach is advantageous, as it ensures that any observed differences in measurements can be attributed primarily to operator-related factors. Nonetheless, it remains important to consider whether machine and probe selection itself may influence reproducibility. Addressing this question in future work would provide valuable insights into the generalisability of these findings and help determine the extent to which equipment variability contributes to measurement reliability.

7.4.3 Conclusion

In conclusion, this study provides the first systematic evaluation of OxNNet-derived placental metrics, demonstrating promising reproducibility in both intra- and inter-operator analyses. By incorporating a rigorous repeated-measures design that accounted for both image acquisition and processing, the findings present a more realistic appraisal of real-world performance than prior segmentation-based studies of reproducibility using VOCAL™. Although limitations remain, particularly regarding the absence of detailed power Doppler gain data, the results highlight the potential of OxNNet parameters and underscore the importance of future work to refine svFMBV measures, to establish an acceptable measurement difference for clinical use, and to validate gain selection reproducibility.

[8] CONCLUSION

8.1 Summary of Key Findings

This thesis presents a comprehensive investigation into the automated assessment of first trimester placental volume and uterine vasculature using three-dimensional power Doppler ultrasound, with the overarching aim of improving the prediction of adverse perinatal outcomes, specifically FGR and PE. Through the refinement and evaluation of the OxNNet deep learning framework, it has been demonstrated that fully automated, reproducible, and rapid segmentation of the placenta and quantification of single vessel fractional moving blood volume (svFMBV) are achievable in a clinical context.

The work is underpinned by two large, prospectively collected cohorts — FirstPLUS and OxPLUS — enabling model development in one cohort and external assessment in the other, across different clinical environments.

The establishment of novel, prescriptive centile charts for first trimester placental volume and svFMBV at multiple anatomical sites represents a novel contribution. These reference ranges, derived from a large, healthy population, showed good apparent fit within the cohort in which they were constructed, but they have not yet been applied to an independent population, so their performance elsewhere remains unknown. Z-scores calculated from these charts were found to be significantly associated with FGR and PE, supporting their potential utility as early biomarkers.

Multivariable predictive models incorporating OxNNet-derived placental metrics, maternal history, and established serum biomarkers achieved good discrimination for both FGR and PE. Discrimination was attenuated but preserved after optimism correction and on assessment in an independent cohort, although this represents a single external assessment rather than established validity. Notably, models constructed using only OxNNet-derived metrics performed moderately well, suggesting that automated placental assessment could be valuable even in settings where access to detailed clinical or biochemical data is limited.

This thesis also introduces, on an exploratory basis, a suite of candidate quantitative markers of placental and vascular morphology, including UPI surface area, SVR, and spiral artery jet

characteristics. Several of these markers demonstrated significant associations with FGR and, to a lesser extent, PE, offering new insights into the pathophysiology of placental insufficiency syndromes. The feasibility of training NHS sonographers in advanced 3D placental ultrasound was established, and the implementation of structured feedback and quality control tools, including CUSUM analysis, was shown to be a viable means of performance monitoring. Most OxNNet-derived metrics exhibited good intra-operator and moderate inter-operator reproducibility, supporting their reliability for research use, although reproducibility would need to be established prospectively before any clinical application.

8.2 Strengths and Limitations

A principal strength of this thesis lies in the scale and quality of the datasets, which enabled both the development of predictive models and their assessment in an independent cohort. The integration of advanced machine learning for automated image analysis represents a methodological advance, facilitating reproducibility and scalability. The statistical approach, which included assessment of discrimination in an independent cohort and of calibration within the development cohort, further enhances the credibility of the findings, although this represents a single external assessment rather than established validity.

However, several limitations must be acknowledged. Processing failures, particularly in cases with higher maternal BMI or certain ethnic backgrounds, may introduce bias and limit the generalisability of the findings. The study was conducted at UK centres using specific ultrasound equipment, and further multicentre and multi-platform validation is required to confirm the broader applicability of the results. The relatively low number of preterm events may limit the performance of predictive models for the most severe cases of FGR and PE. Additionally, the lack of detailed adjustment for aspirin exposure, a known modifier of risk, may attenuate observed associations between imaging markers and outcomes.

8.3 Areas for Future Development

The work presented in this thesis opens several avenues for future research. Multicentre and international validation studies are needed to ensure the generalisability of OxNNet-derived metrics and predictive models across diverse populations, healthcare settings, and ultrasound platforms. Further refinement and anatomical validation of novel vascular and morphological markers are warranted, particularly to reduce processing failures and clarify their pathophysiological significance.

Prospective studies should evaluate the integration of automated placental assessment into clinical workflows, including its impact on decision-making, resource allocation, and perinatal outcomes. The systematic investigation of machine and probe effects on measurement reproducibility and predictive accuracy is also necessary. Embedding real-time feedback and quality assurance tools within clinical ultrasound systems could further enhance measurement reliability and operator performance.

Finally, the novel markers identified in this work provide a foundation for studies aimed at elucidating the biological processes underlying placental insufficiency syndromes, with the potential to inform the development of new preventive or therapeutic strategies.

8.4 Conclusion

In conclusion, this thesis demonstrates that automated, quantitative assessment of the first trimester placenta and uterine vasculature is feasible, and provides proof of concept that the resulting measures carry information about the risk of FGR and PE. The integration of advanced imaging biomarkers into predictive models holds promise for improving the early identification of pregnancies at risk of FGR and PE, thereby supporting more personalised and effective antenatal care. Continued refinement, validation, and clinical integration of these tools will be essential to realise their full potential in improving maternal and perinatal outcomes. The work presented here establishes a foundation for future research in the field of placental imaging and risk prediction.

[9] REFERENCES

1. De Witt F. An historical study on theories of the placenta to 1900. *J Hist Med Allied Sci* 1959;14:360-74. DOI: 10.1093/jhmas/xiv.7.360.
2. Redman CW, Sargent IL. Latest advances in understanding preeclampsia. *Science* 2005;308(5728):1592-4. DOI: 10.1126/science.1111726.
3. Burton GJ, Redman CW, Roberts JM, Moffett A. Pre-eclampsia: pathophysiology and clinical implications. *Bmj* 2019;366:l2381. DOI: 10.1136/bmj.l2381.
4. Flenady V, Koopmans L, Middleton P, et al. Major risk factors for stillbirth in high-income countries: a systematic review and meta-analysis. *Lancet* 2011;377(9774):1331-40. DOI: 10.1016/S0140-6736(10)62233-7.
5. Gardosi J, Madurasinghe V, Williams M, Malik A, Francis A. Maternal and fetal risk factors for stillbirth: population based study. *BMJ* 2013;346:f108. DOI: 10.1136/bmj.f108.
6. Lawn JE, Blencowe H, Waiswa P, et al. Stillbirths: rates, risk factors, and acceleration towards 2030. *Lancet* 2016;387(10018):587-603. DOI: 10.1016/S0140-6736(15)00837-5.
7. Cresswell JA, Alexander M, Chong MYC, et al. Global and regional causes of maternal deaths 2009-20: a WHO systematic analysis. *Lancet Glob Health* 2025;13(4):e626-e634. DOI: 10.1016/s2214-109x(24)00560-6.
8. Gordijn SJ, Beune IM, Thilaganathan B, et al. Consensus definition of fetal growth restriction: a Delphi procedure. *Ultrasound Obstet Gynecol* 2016;48(3):333-9. DOI: 10.1002/uog.15884.
9. Lindqvist PG, Molin J. Does antenatal identification of small-for-gestational age fetuses significantly improve their outcome? *Ultrasound Obstet Gynecol* 2005;25(3):258-64. DOI: 10.1002/uog.1806.
10. Saving Babies' Lives Version Three. A Care Bundle for Reducing Perinatal Mortality. Leeds: NHS England; 2023.
11. Dombrowski MP, Schatz M. Asthma in pregnancy. *Clin Obstet Gynecol* 2010;53(2):301-10. DOI: 10.1097/GRF.0b013e3181de8906.
12. Dadhwal V, Sharma AK, Deka D, Gupta B, Mittal S. The obstetric outcome following treatment in a cohort of patients with antiphospholipid antibody syndrome in a tertiary care center. *J Postgrad Med* 2011;57(1):16-9. DOI: 10.4103/0022-3859.74285.
13. Gelson E, Curry R, Gatzoulis MA, et al. Effect of maternal heart disease on fetal growth. *Obstet Gynecol* 2011;117(4):886-891. DOI: 10.1097/AOG.0b013e31820cab69.
14. Bayliss H, Churchill D, Beevers M, Beevers DG. Anti-hypertensive drugs in pregnancy and fetal growth: evidence for "pharmacological programming" in the first trimester? *Hypertens Pregnancy* 2002;21(2):161-74. DOI: 10.1081/prg-120013785.
15. Pinto SM, Dodd S, Walkinshaw SA, Siney C, Kakkar P, Mousa HA. Substance abuse during pregnancy: effect on pregnancy outcomes. *Eur J Obstet Gynecol Reprod Biol* 2010;150(2):137-41. DOI: 10.1016/j.ejogrb.2010.02.026.

16. Belkacemi L, Nelson DM, Desai M, Ross MG. Maternal undernutrition influences placental-fetal development. *Biol Reprod* 2010;83(3):325-31. DOI: 10.1095/biolreprod.110.084517.
17. Usman S, Lawin-O'Brien A, Lees C. Differential Diagnosis of Fetal Growth Restriction. In: Lees C, Visser GHA, Hecher K, eds. *Placental-Fetal Growth Restriction*. Cambridge: Cambridge University Press; 2018:14-30.
18. Snijders RJ, Sherrod C, Gosden CM, Nicolaides KH. Fetal growth retardation: associated malformations and chromosomal abnormalities. *Am J Obstet Gynecol* 1993;168(2):547-55. DOI: 10.1016/0002-9378(93)90491-z.
19. Wilkins-Haug L, Roberts DJ, Morton CC. Confined placental mosaicism and intrauterine growth retardation: a case-control analysis of placentas at delivery. *Am J Obstet Gynecol* 1995;172(1 Pt 1):44-50. DOI: 10.1016/0002-9378(95)90082-9.
20. Boghossian NS, Horbar JD, Murray JC, Carpenter JH. Anthropometric charts for infants with trisomies 21, 18, or 13 born between 22 weeks gestation and term: the VON charts. *Am J Med Genet A* 2012;158a(2):322-32. DOI: 10.1002/ajmg.a.34423.
21. Rosenthal GL, Wilson PD, Permutt T, Boughman JA, Ferencz C. Birth weight and cardiovascular malformations: a population-based study. The Baltimore-Washington Infant Study. *Am J Epidemiol* 1991;133(12):1273-81. DOI: 10.1093/oxfordjournals.aje.a115839.
22. Rosenthal GL. Patterns of prenatal growth among infants with cardiovascular malformations: possible fetal hemodynamic effects. *Am J Epidemiol* 1996;143(5):505-13. DOI: 10.1093/oxfordjournals.aje.a008771.
23. Hussain U, Daemen A, Missfelder-Lobos H, et al. Umbilical artery pulsatility index and fetal abdominal circumference in isolated gastroschisis. *Ultrasound Obstet Gynecol* 2011;38(5):538-42. DOI: 10.1002/uog.8947.
24. Norman SM, Odibo AO, Longman RE, Roehl KA, Macones GA, Cahill AG. Neural tube defects and associated low birth weight. *Am J Perinatol* 2012;29(6):473-6. DOI: 10.1055/s-0032-1304830.
25. Sankaran S, Kyle PM. Aetiology and pathogenesis of IUGR. *Best Pract Res Clin Obstet Gynaecol* 2009;23(6):765-77. DOI: 10.1016/j.bpobgyn.2009.05.003.
26. Dahlstrøm BL, Engh ME, Bukholm G, Oian P. Changes in the prevalence of pre-eclampsia in Akershus County and the rest of Norway during the past 35 years. *Acta Obstet Gynecol Scand* 2006;85(8):916-21. DOI: 10.1080/00016340500442449.
27. Klungsøyr K, Morken NH, Irgens L, Vollset SE, Skjaerven R. Secular trends in the epidemiology of pre-eclampsia throughout 40 years in Norway: prevalence, risk factors and perinatal survival. *Paediatr Perinat Epidemiol* 2012;26(3):190-8. DOI: 10.1111/j.1365-3016.2012.01260.x.
28. Sun S, Li W, Zhang X, Aziz AuR, Zhang N. Trends in global and regional incidence and prevalence of hypertensive disorders in pregnancy (1990–2021): an age-period-cohort analysis. *Scientific Reports* 2025;15(1):1513. DOI: 10.1038/s41598-025-85819-0.
29. Gestational Hypertension and Preeclampsia: ACOG Practice Bulletin, Number 222. *Obstet Gynecol* 2020;135(6):e237-e260. DOI: 10.1097/AOG.0000000000003891.

30. Magee LA, Brown MA, Hall DR, et al. The 2021 International Society for the Study of Hypertension in Pregnancy classification, diagnosis & management recommendations for international practice. *Pregnancy Hypertens* 2022;27:148-169. DOI: 10.1016/j.preghy.2021.09.008.
31. National Institute for Health and Care Excellence. Hypertension in pregnancy: diagnosis and management (NG133). 2019.
32. Tassi A, Sala A, Mazzera I, Restaino S, Vizzielli G, Driul L. Long-term outcomes of patients with preeclampsia, a review of the literature. *Hypertens Pregnancy* 2023;42(1):2217448. DOI: 10.1080/10641955.2023.2217448.
33. Haßdenteufel K, Müller M, Gutsfeld R, et al. Long-term effects of preeclampsia on maternal cardiovascular health and postpartum utilization of primary care: an observational claims data study. *Archives of Gynecology and Obstetrics* 2023;307(1):275-284.
34. Brown MC, Best KE, Pearce MS, Waugh J, Robson SC, Bell R. Cardiovascular disease risk in women with pre-eclampsia: systematic review and meta-analysis. *Eur J Epidemiol* 2013;28(1):1-19. DOI: 10.1007/s10654-013-9762-6.
35. Melchiorre K, Thilaganathan B, Giorgione V, Ridder A, Memmo A, Khalil A. Hypertensive Disorders of Pregnancy and Future Cardiovascular Health. *Front Cardiovasc Med* 2020;7:59. DOI: 10.3389/fcvm.2020.00059.
36. Lykke JA, Langhoff-Roos J, Sibai BM, Funai EF, Triche EW, Paidas MJ. Hypertensive pregnancy disorders and subsequent cardiovascular morbidity and type 2 diabetes mellitus in the mother. *Hypertension* 2009;53(6):944-51. DOI: 10.1161/hypertensionaha.109.130765.
37. Ngoc NT, Merialdi M, Abdel-Aleem H, et al. Causes of stillbirths and early neonatal deaths: data from 7993 pregnancies in six developing countries. *Bull World Health Organ* 2006;84(9):699-705. DOI: 10.2471/blt.05.027300.
38. Aughey H, Jardine J, Knight H, et al. Iatrogenic and spontaneous preterm birth in England: A population-based cohort study. *BJOG* 2023;130(1):33-41. DOI: 10.1111/1471-0528.17291.
39. Redman CW, Sargent IL, Staff AC. IFPA Senior Award Lecture: making sense of pre-eclampsia - two placental causes of preeclampsia? *Placenta* 2014;35 Suppl:S20-5. DOI: 10.1016/j.placenta.2013.12.008.
40. Ganzevoort JW, Thilaganathan B. Definition of Fetal Growth Restriction and Uteroplacental Insufficiency. In: Lees C, Visser GHA, Hecher K, eds. *Placental-Fetal Growth Restriction*. Cambridge: Cambridge University Press; 2018:6-13.
41. Pijnenborg R, Vercruysse L, Hanssens M. The uterine spiral arteries in human pregnancy: facts and controversies. *Placenta* 2006;27(9-10):939-58. DOI: 10.1016/j.placenta.2005.12.006.
42. Khong Y, Brosens I. Defective deep placentation. *Best Pract Res Clin Obstet Gynaecol* 2011;25(3):301-11. DOI: 10.1016/j.bpobgyn.2010.10.012.
43. Pijnenborg R, Bland JM, Robertson WB, Brosens I. Uteroplacental arterial changes related to interstitial trophoblast migration in early human pregnancy. *Placenta* 1983;4(4):397-413. DOI: 10.1016/s0143-4004(83)80043-5.

44. Robertson WB, Khong TY, Brosens I, De Wolf F, Sheppard BL, Bonnar J. The placental bed biopsy: review from three European centers. *Am J Obstet Gynecol* 1986;155(2):401-12. DOI: 10.1016/0002-9378(86)90843-4.
45. Lyall F. Development of the utero-placental circulation: the role of carbon monoxide and nitric oxide in trophoblast invasion and spiral artery transformation. *Microsc Res Tech* 2003;60(4):402-11. DOI: 10.1002/jemt.10278.
46. Lyall F. The human placental bed revisited. *Placenta* 2002;23(8-9):555-62. DOI: 10.1053/plac.2002.0850.
47. Abstracts to be presented at the Conference of the International Federation of Placenta Associations, 9th Meeting of the European Placenta Group. *Placenta* 2001;22(7):A1-A80. DOI: <https://doi.org/10.1053/plac.2001.0712>.
48. Robson S, Ball E, Lyall F, Simpson H, Ayis H, Bulmer J. Endovascular trophoblast invasion and spiral artery transformation: the “two wave” theory revisited. *Placenta* 2001;22:A25.
49. Brosens I, Dixon HG, Robertson WB. Fetal growth retardation and the arteries of the placental bed. *Br J Obstet Gynaecol* 1977;84(9):656-63. DOI: 10.1111/j.1471-0528.1977.tb12676.x.
50. De Wolf F, Brosens I, Renaer M. Fetal growth retardation and the maternal arterial supply of the human placenta in the absence of sustained hypertension. *Br J Obstet Gynaecol* 1980;87(8):678-85. DOI: 10.1111/j.1471-0528.1980.tb04601.x.
51. Khong TY, De Wolf F, Robertson WB, Brosens I. Inadequate maternal vascular response to placentation in pregnancies complicated by pre-eclampsia and by small-for-gestational age infants. *Br J Obstet Gynaecol* 1986;93(10):1049-59. DOI: 10.1111/j.1471-0528.1986.tb07830.x.
52. Brosens JJ, Pijnenborg R, Brosens IA. The myometrial junctional zone spiral arteries in normal and abnormal pregnancies: a review of the literature. *Am J Obstet Gynecol* 2002;187(5):1416-23. DOI: 10.1067/mob.2002.127305.
53. Burton GJ, Woods AW, Jauniaux E, Kingdom JC. Rheological and physiological consequences of conversion of the maternal spiral arteries for uteroplacental blood flow during human pregnancy. *Placenta* 2009;30(6):473-82. DOI: 10.1016/j.placenta.2009.02.009.
54. Redman CW. Preeclampsia: a multi-stress disorder. *Rev Med Interne* 2011;32 Suppl 1:S41-4. DOI: 10.1016/j.revmed.2011.03.331.
55. Burton GJ, Yung HW, Cindrova-Davies T, Charnock-Jones DS. Placental endoplasmic reticulum stress and oxidative stress in the pathophysiology of unexplained intrauterine growth restriction and early onset preeclampsia. *Placenta* 2009;30 Suppl A(Suppl):S43-8. DOI: 10.1016/j.placenta.2008.11.003.
56. Sheppard BL, Bonnar J. The ultrastructure of the arterial supply of the human placenta in pregnancy complicated by fetal growth retardation. *Br J Obstet Gynaecol* 1976;83(12):948-59. DOI: 10.1111/j.1471-0528.1976.tb00781.x.
57. Pijnenborg R, Anthony J, Davey DA, et al. Placental bed spiral arteries in the hypertensive disorders of pregnancy. *Br J Obstet Gynaecol* 1991;98(7):648-55. DOI: 10.1111/j.1471-0528.1991.tb13450.x.

58. Hung TH, Skepper JN, Burton GJ. In vitro ischemia-reperfusion injury in term human placenta as a model for oxidative stress in pathological pregnancies. *Am J Pathol* 2001;159(3):1031-43. DOI: 10.1016/s0002-9440(10)61778-6.
59. Jauniaux E, Ramsay B, Campbell S. Ultrasonographic investigation of placental morphologic characteristics and size during the second trimester of pregnancy. *Am J Obstet Gynecol* 1994;170(1 Pt 1):130-7. DOI: 10.1016/s0002-9378(94)70397-3.
60. Robertson WB. Uteroplacental vasculature. *Journal of Clinical Pathology* 1976;s3-10(1):9-17. DOI: 10.1136/jcp.s3-10.1.9.
61. Collins SL, Birks JS, Stevenson GN, Papageorgiou AT, Noble JA, Impey L. Measurement of spiral artery jets: general principles and differences observed in small-for-gestational-age pregnancies. *Ultrasound Obstet Gynecol* 2012;40(2):171-8. DOI: 10.1002/uog.10149.
62. Collins SL, Stevenson GN, Noble JA, Impey L. Developmental changes in spiral artery blood flow in the human placenta observed with colour Doppler ultrasonography. *Placenta* 2012;33(10):782-7. DOI: 10.1016/j.placenta.2012.07.005.
63. Saghian R, James JL, Tawhai MH, Collins SL, Clark AR. Association of Placental Jets and Mega-Jets With Reduced Villous Density. *J Biomech Eng* 2017;139(5):0510011-05100110. DOI: 10.1115/1.4036145.
64. Pedroso MA, Palmer KR, Hodges RJ, Costa FDS, Rolnik DL. Uterine Artery Doppler in Screening for Preeclampsia and Fetal Growth Restriction. *Rev Bras Ginecol Obstet* 2018;40(5):287-293. DOI: 10.1055/s-0038-1660777.
65. Poon LC, Syngelaki A, Akolekar R, Lai J, Nicolaides KH. Combined screening for preeclampsia and small for gestational age at 11-13 weeks. *Fetal Diagn Ther* 2013;33(1):16-27. DOI: 10.1159/000341712.
66. Akolekar R, Syngelaki A, Poon L, Wright D, Nicolaides KH. Competing risks model in early screening for preeclampsia by biophysical and biochemical markers. *Fetal Diagn Ther* 2013;33(1):8-15. DOI: 10.1159/000341264.
67. Schaaps JP, Tsatsaris V, Goffin F, et al. Shunting the intervillous space: new concepts in human uteroplacental vascularization. *Am J Obstet Gynecol* 2005;192(1):323-32. DOI: 10.1016/j.ajog.2004.06.066.
68. Heckel GP, Tobin CE. Arteriovenous shunts in the myometrium. *Am J Obstet Gynecol* 1956;71(1):199-205. DOI: 10.1016/0002-9378(56)90695-0.
69. Clark AR, James JL, Stevenson GN, Collins SL. Understanding abnormal uterine artery Doppler waveforms: A novel computational model to explore potential causes within the utero-placental vasculature. *Placenta* 2018;66:74-81. DOI: 10.1016/j.placenta.2018.05.001.
70. Steer C, Campbell S, Pampiglione J, Kingsland C, Mason B, Collins W. Transvaginal colour flow imaging of the uterine arteries during the ovarian and menstrual cycles. *Human Reproduction* 1990;5(4):391-395.
71. Tan, Zaidi, Campbell, Doyle, Collins. Blood flow changes in the ovarian and uterine arteries during the normal menstrual cycle. *American journal of obstetrics and gynecology* 1996;175(3P1):625-631.

72. Scholtes MC, Wladimiroff JW, van Rijen HJ, Hop WC. Uterine and ovarian flow velocity waveforms in the normal menstrual cycle: a transvaginal Doppler study. *Fertility and sterility* 1989;52(6):981-985.
73. Järvelä I, Tekay A, Jouppila P. The effect of diclofenac on uterine artery blood flow resistance during menstruation in patients with and without a copper intrauterine device. *Human reproduction (Oxford, England)* 1998;13(9):2480-2483.
74. Hale SA, Jones CW, Osol G, Schonberg A, Badger GJ, Bernstein IM. Sildenafil increases uterine blood flow in nonpregnant nulliparous women. *Reproductive Sciences* 2010;17(4):358-365.
75. Järvelä I, Tekay A, Jouppila P. The effect of a levonorgestrel-releasing intrauterine system on uterine artery blood flow, hormone concentrations and ovarian cyst formation in fertile women. *Human Reproduction (Oxford, England)* 1998;13(12):3379-3383.
76. Jiménez MF, Arbo E, Vettori D, de Freitas FM, Cunha-Filho JSL. The effect of the levonorgestrel-releasing intrauterine system and the copper intrauterine device on subendometrial microvascularization and uterine artery blood flow. *Fertility and sterility* 2008;90(5):1574-1578.
77. Collins S, Grant D, Black R, Vellayan M, Impey L. Abdominal pregnancy: a perfusion confusion? *Placenta* 2011;32(10):793-795.
78. Morris RK, Johnstone E, Lees C, Morton V, Smith G. Investigation and Care of a Small-for-Gestational-Age Fetus and a Growth Restricted Fetus (Green-top Guideline No. 31). *BJOG* 2024;131(9):e31-e80. DOI: 10.1111/1471-0528.17814.
79. Fetal Growth Restriction: ACOG Practice Bulletin, Number 227. *Obstet Gynecol* 2021;137(2):e16-e28. DOI: 10.1097/aog.0000000000004251.
80. Melamed N, Baschat A, Yinon Y, et al. FIGO (international Federation of Gynecology and obstetrics) initiative on fetal growth: best practice advice for screening, diagnosis, and management of fetal growth restriction. *Int J Gynaecol Obstet* 2021;152 Suppl 1(Suppl 1):3-57. DOI: 10.1002/ijgo.13522.
81. Poon LC, Shennan A, Hyett JA, et al. The International Federation of Gynecology and Obstetrics (FIGO) initiative on pre-eclampsia: A pragmatic guide for first-trimester screening and prevention. *Int J Gynaecol Obstet* 2019;145 Suppl 1(Suppl 1):1-33. DOI: 10.1002/ijgo.12802.
82. Rolnik DL, Wright D, Poon LC, et al. Aspirin versus Placebo in Pregnancies at High Risk for Preterm Preeclampsia. *N Engl J Med* 2017;377(7):613-622. DOI: 10.1056/NEJMoa1704559.
83. Bujold E, Roberge S, Lacasse Y, et al. Prevention of preeclampsia and intrauterine growth restriction with aspirin started in early pregnancy: a meta-analysis. *Obstet Gynecol* 2010;116(2 Pt 1):402-414. DOI: 10.1097/AOG.0b013e3181e9322a.
84. Roberge S, Villa P, Nicolaides K, et al. Early administration of low-dose aspirin for the prevention of preterm and term preeclampsia: a systematic review and meta-analysis. *Fetal Diagn Ther* 2012;31(3):141-6. DOI: 10.1159/000336662.
85. Poon LC, Wright D, Thornton S, Akolekar R, Brocklehurst P, Nicolaides KH. Efficacy and Mechanism Evaluation. Mini-combined test compared with NICE guidelines for early

risk-assessment for pre-eclampsia: the SPREE diagnostic accuracy study. Southampton (UK): NIHR Journals Library; 2020.

86. Tan MY, Koutoulas L, Wright D, Nicolaides KH, Poon LCY. Protocol for the prospective validation study: 'Screening programme for pre-eclampsia' (SPREE). *Ultrasound Obstet Gynecol* 2017;50(2):175-179. DOI: 10.1002/uog.17467.
87. Tan MY, Wright D, Syngelaki A, et al. Comparison of diagnostic accuracy of early screening for pre-eclampsia by NICE guidelines and a method combining maternal factors and biomarkers: results of SPREE. *Ultrasound Obstet Gynecol* 2018;51(6):743-750. DOI: 10.1002/uog.19039.
88. Crovetto F, Crispi F, Scazzocchio E, et al. First-trimester screening for early and late small-for-gestational-age neonates using maternal serum biochemistry, blood pressure and uterine artery Doppler. *Ultrasound Obstet Gynecol* 2014;43(1):34-40. DOI: 10.1002/uog.12537.
89. Crovetto F, Triunfo S, Crispi F, et al. Differential performance of first-trimester screening in predicting small-for-gestational-age neonate or fetal growth restriction. *Ultrasound Obstet Gynecol* 2017;49(3):349-356. DOI: 10.1002/uog.15919.
90. Crovetto F, Triunfo S, Crispi F, et al. First-trimester screening with specific algorithms for early- and late-onset fetal growth restriction. *Ultrasound Obstet Gynecol* 2016;48(3):340-8. DOI: 10.1002/uog.15879.
91. McCowan LM, Thompson JM, Taylor RS, et al. Prediction of Small for Gestational Age Infants in Healthy Nulliparous Women Using Clinical and Ultrasound Risk Factors Combined with Early Pregnancy Biomarkers. *PLoS One* 2017;12(1):e0169311. DOI: 10.1371/journal.pone.0169311.
92. Papastefanou I, Nowacka U, Syngelaki A, et al. Competing risks model for prediction of small-for-gestational-age neonates from biophysical markers at 19 to 24 weeks' gestation. *Am J Obstet Gynecol* 2021;225(5):530.e1-530.e19. DOI: 10.1016/j.ajog.2021.04.247.
93. Papastefanou I, Wright D, Nicolaides KH. Competing-risks model for prediction of small-for-gestational-age neonate from maternal characteristics and medical history. *Ultrasound Obstet Gynecol* 2020;56(2):196-205. DOI: 10.1002/uog.22129.
94. Karagiannis G, Akolekar R, Sarquis R, Wright D, Nicolaides KH. Prediction of small-for-gestation neonates from biophysical and biochemical markers at 11-13 weeks. *Fetal Diagn Ther* 2011;29(2):148-54. DOI: 10.1159/000321694.
95. Roberge S, Nicolaides K, Demers S, Hyett J, Chaillet N, Bujold E. The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. *Am J Obstet Gynecol* 2017;216(2):110-120 e6. DOI: 10.1016/j.ajog.2016.09.076.
96. Little WA. The significance of placental/fetal weight ratios. *Am J Obstet Gynecol* 1960;79:134-7. DOI: 10.1016/0002-9378(60)90372-0.
97. Sinclair JG. Placental-fetal weight ratios. *Tex Rep Biol Med* 1948;6(2):168-75.
98. Molteni RA. Placental growth and fetal/placental weight (F/P) ratios throughout gestation--their relationship to patterns of fetal growth. *Semin Perinatol* 1984;8(2):94-100.

99. Hellman LM, Kobayashi M, Tolles WE, Cromb E. Ultrasonic studies on the volumetric growth of the human placenta. *Am J Obstet Gynecol* 1970;108(5):740-50. DOI: 10.1016/0002-9378(70)90540-5.
100. Bleker OP, Kloosterman GJ, Breur W, Mieras DJ. The volumetric growth of the human placenta: a longitudinal ultrasonic study. *Am J Obstet Gynecol* 1977;127(6):657-61. DOI: 10.1016/0002-9378(77)90369-6.
101. Jones TB, Price RR, Gibbs SJ. Volumetric determination of placental and uterine growth relationships from B-mode ultrasound by serial area-volume determinations. *Invest Radiol* 1981;16(2):101-6. DOI: 10.1097/00004424-198103000-00005.
102. Law LW, Leung TY, Sahota DS, Chan LW, Fung TY, Lau TK. Which ultrasound or biochemical markers are independent predictors of small-for-gestational age? *Ultrasound Obstet Gynecol* 2009;34(3):283-7. DOI: 10.1002/uog.6455.
103. Hafner E, Metzenbauer M, Dillinger-Paller B, et al. Correlation of first trimester placental volume and second trimester uterine artery Doppler flow. *Placenta* 2001;22(8-9):729-34. DOI: 10.1053/plac.2001.0721.
104. Plasencia W, Akolekar R, Dagklis T, Veduta A, Nicolaides KH. Placental volume at 11-13 weeks' gestation in the prediction of birth weight percentile. *Fetal Diagn Ther* 2011;30(1):23-8. DOI: 10.1159/000324318.
105. Hafner E, Metzenbauer M, Hofinger D, et al. Comparison between three-dimensional placental volume at 12 weeks and uterine artery impedance/notching at 22 weeks in screening for pregnancy-induced hypertension, pre-eclampsia and fetal growth restriction in a low-risk population. *Ultrasound Obstet Gynecol* 2006;27(6):652-7. DOI: 10.1002/uog.2641.
106. Stevenson GN, Collins SL, Ding J, Impey L, Noble JA. 3-D Ultrasound Segmentation of the Placenta Using the Random Walker Algorithm: Reliability and Agreement. *Ultrasound Med Biol* 2015;41(12):3182-93. DOI: 10.1016/j.ultrasmedbio.2015.07.021.
107. Aye CY, Stevenson GN, Impey L, Collins SL. Comparison of 2-D and 3-D estimates of placental volume in early pregnancy. *Ultrasound Med Biol* 2015;41(3):734-40. DOI: 10.1016/j.ultrasmedbio.2014.10.001.
108. Cheong KB, Leung KY, Li TK, Chan HY, Lee YP, Tang MH. Comparison of inter- and intraobserver agreement and reliability between three different types of placental volume measurement technique (XI VOCAL, VOCAL and multiplanar) and validity in the in-vitro setting. *Ultrasound Obstet Gynecol* 2010;36(2):210-7. DOI: 10.1002/uog.7609.
109. Grady L. Random walks for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 2006;28(11):1768-1783.
110. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436-44. DOI: 10.1038/nature14539.
111. Looney P, Stevenson GN, Nicolaides KH, et al. Fully automated, real-time 3D ultrasound segmentation to estimate first trimester placental volume using deep learning. *JCI Insight* 2018;3(11). DOI: 10.1172/jci.insight.120178.
112. Looney P, Yin Y, Collins SL, et al. Fully Automated 3-D Ultrasound Segmentation of the Placenta, Amniotic Fluid, and Fetus for Early Pregnancy Assessment. *IEEE Trans Ultrason Ferroelectr Freq Control* 2021;68(6):2038-2047. DOI: 10.1109/tuffc.2021.3052143.

113. Looney P, Stevenson G, Nicolaides K, et al. Prediction of term small for gestational age babies using first trimester placental volume; a comparison of a novel, fully automated technique, Oxnnet with a commercially available, semi-automatic tool, VOCAL (tm). *Placenta* 2018;69.
114. Welsh A. Quantification of power Doppler and the index 'fractional moving blood volume' (FMBV). *Ultrasound Obstet Gynecol* 2004;23(4):323-6. DOI: 10.1002/uog.1037.
115. Gudmundsson S, Valentin L, Pirhonen J, Olofsson PA, Dubiel M, Marsal K. Factors affecting color Doppler energy ultrasound recordings in an in-vitro model. *Ultrasound Med Biol* 1998;24(6):899-902. DOI: 10.1016/s0301-5629(98)00031-3.
116. Preidler KW, Szolar DM, Uggowitz M, Stiskal M, Horina J. Technical note: comparison of colour Doppler energy sonography with conventional colour Doppler sonography in detection of flow signal in peripheral renal transplant vessels. *Br J Radiol* 1995;68(814):1103-5. DOI: 10.1259/0007-1285-68-814-1103.
117. Pairleitner H, Steiner H, Hasenoehrl G, Staudach A. Three-dimensional power Doppler sonography: imaging and quantifying blood flow and vascularization. *Ultrasound Obstet Gynecol* 1999;14(2):139-43. DOI: 10.1046/j.1469-0705.1999.14020139.x.
118. Welsh AW, Collins SL, Stevenson GN, Noble JA, Impey L. Inapplicability of fractional moving blood volume technique to standardize Virtual Organ Computer-aided Analysis indices for quantified three-dimensional power Doppler. *Ultrasound Obstet Gynecol* 2012;40(6):688-92. DOI: 10.1002/uog.11139.
119. Jones N, Deshpande R, Mousa H, Mansell P, Raine-Fenning N, Bugg G. Fractional volume of placental vessels in women with diabetes using a novel stereological 3D power Doppler technique. *Placenta* 2013;34(11):1002-1008.
120. Collins SL, Stevenson GN, Noble JA, Impey L, Welsh AW. Influence of power Doppler gain setting on Virtual Organ Computer-aided Analysis indices in vivo: can use of the individual sub-noise gain level optimize information? *Ultrasound Obstet Gynecol* 2012;40(1):75-80. DOI: 10.1002/uog.10122.
121. Rubin JM, Adler RS, Fowlkes JB, et al. Fractional moving blood volume: estimation with power Doppler US. *Radiology* 1995;197(1):183-90. DOI: 10.1148/radiology.197.1.7568820.
122. Rubin JM, Bude RO, Fowlkes JB, Spratt RS, Carson PL, Adler RS. Normalizing fractional moving blood volume estimates with power Doppler US: defining a stable intravascular point with the cumulative power distribution function. *Radiology* 1997;205(3):757-65. DOI: 10.1148/radiology.205.3.9393532.
123. Welsh AW, Fowlkes JB, Pinter SZ, et al. Three-dimensional US Fractional Moving Blood Volume: Validation of Renal Perfusion Quantification. *Radiology* 2019;293(2):460-468. DOI: 10.1148/radiol.2019190248.
124. Yin Y, Clark AR, Collins SL. 3D Single Vessel Fractional Moving Blood Volume (3D-svFMBV): Fully Automated Tissue Perfusion Estimation Using Ultrasound. *IEEE Trans Med Imaging* 2024;43(7):2707-2717. DOI: 10.1109/tmi.2024.3376668.
125. Collins SL, Welsh AW, Impey L, Noble JA, Stevenson GN. 3D fractional moving blood volume (3D-FMBV) demonstrates decreased first trimester placental vascularity in pre-

eclampsia but not the term, small for gestation age baby. PLoS One 2017;12(6):e0178675. DOI: 10.1371/journal.pone.0178675.

126. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370(9596):1453-7. DOI: 10.1016/S0140-6736(07)61602-X.
127. Ciobanu A, Wright A, Syngelaki A, Wright D, Akolekar R, Nicolaides KH. Fetal Medicine Foundation reference ranges for umbilical artery and middle cerebral artery pulsatility index and cerebroplacental ratio. Ultrasound Obstet Gynecol 2019;53(4):465-472. DOI: 10.1002/uog.20157.
128. Nicolaides KH, Wright D, Syngelaki A, Wright A, Akolekar R. Fetal Medicine Foundation fetal and neonatal population weight charts. Ultrasound Obstet Gynecol 2018;52(1):44-51. DOI: 10.1002/uog.19073.
129. Diabetes in pregnancy: management from preconception to the postnatal period. London: National Institute for Health and Care Excellence (NICE)

Copyright © NICE 2020.; 2020.

130. Robinson HP, Fleming JE. A critical evaluation of sonar "crown-rump length" measurements. Br J Obstet Gynaecol 1975;82(9):702-10. DOI: 10.1111/j.1471-0528.1975.tb00710.x.
131. Sanderson J, Wu L, Mahajan A, Meriki N, Henry A, Welsh AW. Selection of the sub-noise gain level for acquisition of VOCAL data sets: a reliability study. Ultrasound Med Biol 2014;40(3):562-7. DOI: 10.1016/j.ultrasmedbio.2013.09.004.
132. Looney PT, Stevenson GN, Collins SL. 3D ultrasound file reading and coordinate transformations. J Open Source Softw 2019;4(33) DOI: 10.21105/joss.01063.
133. Ensor J. pmsampsize: Sample Size for Development of a Prediction Model. Version 1.1.3. 2023.
134. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 2019;38(7):1276-1296. DOI: 10.1002/sim.7992.
135. Lehmann EL, D'Abrera HJ. Nonparametrics: statistical methods based on ranks: Springer New York, 2006.
136. McHugh ML. The chi-square test of independence. Biochimica medica 2013;23(2):143-149.
137. Sheikh J, Allotey J, Kew T, et al. Effects of race and ethnicity on perinatal outcomes in high-income and upper-middle-income countries: an individual participant data meta-analysis of 2 198 655 pregnancies. The Lancet 2022;400(10368):2049-2062.
138. Arechvo A, Voicu D, Gil MM, Syngelaki A, Akolekar R, Nicolaides KH. Maternal race and pre-eclampsia: Cohort study and systematic review with meta-analysis. BJOG: An International Journal of Obstetrics & Gynaecology 2022;129(12):2082-2093.
139. Jardine J, Walker K, Gurol-Urganci I, et al. Adverse pregnancy outcomes attributable to socioeconomic and ethnic inequalities in England: a national cohort study. Lancet 2021;398(10314):1905-1912. DOI: 10.1016/s0140-6736(21)01595-6.

140. Brand JS, Gaillard R, West J, et al. Associations of maternal quitting, reducing, and continuing smoking during pregnancy with longitudinal fetal growth: Findings from Mendelian randomization and parental negative control studies. *PLoS medicine* 2019;16(11):e1002972.
141. Wei J, Liu C-X, Gong T-T, Wu Q-J, Wu L. Cigarette smoking during pregnancy and preeclampsia risk: a systematic review and meta-analysis of prospective studies. *Oncotarget* 2015;6(41):43667.
142. Wang J, Yang W, Xiao W, Cao S. The association between smoking during pregnancy and hypertensive disorders of pregnancy: A systematic review and meta-analysis. *International Journal of Gynecology & Obstetrics* 2022;157(1):31-41.
143. Duley L, Meher S, Hunter KE, Seidler AL, Askie LM. Antiplatelet agents for preventing pre-eclampsia and its complications. *Cochrane database of systematic reviews* 2019(10).
144. Creanga AA, Catalano PM, Bateman BT. Obesity in pregnancy. *New England Journal of Medicine* 2022;387(3):248-259.
145. Yeh J, Shelton JA. Increasing prepregnancy body mass index: analysis of trends and contributing variables. *Am J Obstet Gynecol* 2005;193(6):1994-8. DOI: 10.1016/j.ajog.2005.05.001.
146. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology* 1991;181(1):129-33. DOI: 10.1148/radiology.181.1.1887021.
147. Kiserud T, Piaggio G, Carroli G, et al. The World Health Organization Fetal Growth Charts: A Multinational Longitudinal Study of Ultrasound Biometric Measurements and Estimated Fetal Weight. *PLoS Med* 2017;14(1):e1002220. DOI: 10.1371/journal.pmed.1002220.
148. Stirnemann J, Villar J, Salomon LJ, et al. International estimated fetal weight standards of the INTERGROWTH-21. *Ultrasound Obstet Gynecol* 2017;49(4):478-486. DOI: 10.1002/uog.17347.
149. Melamed N, Hiersch L, Aviram A, Mei-Dan E, Keating S, Kingdom JC. Diagnostic accuracy of fetal growth charts for placenta-related fetal growth restriction. *Placenta* 2021;105:70-77.
150. Liauw J, Mayer C, Albert A, Fernandez A, Hutcheon JA. Which chart and which cut-point: deciding on the INTERGROWTH, World Health Organization, or Hadlock fetal growth chart. *BMC Pregnancy and Childbirth* 2022;22(1):25.
151. Visser GH, Nicholson WK, Barnea ER, Ramasauskaite D, Nassar AH, FIGO Safe Motherhood NHC. FIGO position paper on reference charts for fetal growth and size at birth: which one to use? *International Journal of Gynecology & Obstetrics* 2021;152(2):148-151.
152. Meengeonthong D, Luewan S, Sirichotiyakul S, Tongsong T. Reference ranges of placental volume measured by virtual organ computer-aided analysis between 10 and 14 weeks of gestation. *J Clin Ultrasound* 2017;45(4):185-191. DOI: 10.1002/jcu.22441.
153. Trilla Solà C, Parra Roca J, Llurba Olivé E. Reference Ranges of 2-Dimensional Placental Biometry and 3-Dimensional Placental Volume between 11 and 14 Weeks of Gestation. *Diagnostics (Basel)* 2024;14(14) DOI: 10.3390/diagnostics14141556.

154. Guyomard A, Macé G, Ferdynus C, Franceschini C, Rousseau T, Sagot P. Reference ranges and distribution of placental volume by 3-dimensional virtual organ computer-aided analysis between 11 weeks and 13 weeks 6 days. *J Ultrasound Med* 2013;32(8):1477-82. DOI: 10.7863/ultra.32.8.1477.
155. Healy MJR. Notes on the statistics of growth standards. *Annals of Human Biology* 1974;1(1):41-46. DOI: 10.1080/03014467400000041.
156. Kumar F, Kemp J, Edwards C, et al. Pregnancy physiology pattern prediction study (4P study): protocol of an observational cohort study collecting vital sign information to inform the development of an accurate centile-based obstetric early warning score. *BMJ Open* 2017;7(9):e016034. DOI: 10.1136/bmjopen-2017-016034.
157. Peterson RA. Finding optimal normalizing transformations via bestNormalize. *The R Journal* 2021;13:310. DOI: 10.32614/RJ-2021-041.
158. Royston P, Altman DG. Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *Journal of the Royal Statistical Society Series C (Applied Statistics)* 1994;43(3):429-467. DOI: 10.2307/2986270.
159. Wei Y, Pere A, Koenker R, He X. Quantile regression methods for reference growth charts. *Stat Med* 2006;25(8):1369-82. DOI: 10.1002/sim.2271.
160. Cole TJ, Green PJ. Smoothing reference centile curves: The lms method and penalized likelihood. *Statistics in Medicine* 1992;11(10):1305-1319. DOI: <https://doi.org/10.1002/sim.4780111005>.
161. Koenker R. quantreg: Quantile regression. Version 6.1. 2025.
162. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society Series C: Applied Statistics* 2005;54(3):507-554.
163. Sakamoto Y, Ishiguro M, Kitagawa G. Akaike Information Criterion Statistics: Springer Netherlands, 1986.
164. Bilardo CM, Chaoui R, Hyett JA, et al. ISUOG Practice Guidelines (updated): performance of 11-14-week ultrasound scan. *Ultrasound Obstet Gynecol* 2023;61(1):127-143. DOI: 10.1002/uog.26106.
165. Executive summary: Workshop on developing an optimal maternal-fetal medicine ultrasound practice, February 7-8, 2023, cosponsored by the Society for Maternal-Fetal Medicine, American College of Obstetricians and Gynecologists, American Institute of Ultrasound in Medicine, American Registry for Diagnostic Medical Sonography, International Society of Ultrasound in Obstetrics and Gynecology, Gottesfeld-Hohler Memorial Foundation, and Perinatal Quality Foundation. *Am J Obstet Gynecol* 2023;229(3):B20-b24. DOI: 10.1016/j.ajog.2023.06.009.
166. ISUOG Education Committee. ISUOG Education Committee recommendations for basic training in obstetric and gynecological ultrasound. *Ultrasound Obstet Gynecol* 2013;43:113-6.
167. World Health Organization. Maternal and fetal assessment update: imaging ultrasound before 24 weeks of pregnancy. 2022.
168. GOV.UK. Fetal anomaly screening programme (FASP) handbook. London: NHS England, 2015.

169. Wright D. Nuchal translucency and crown rump length diagnostic plot self-assessment tool. NHS England. (<https://www.gov.uk/government/publications/fetal-anomaly-screening-programme-handbook/quality-assurance#departmental-review-of-ultrasound-images>).
170. Balsyte D, Schäffer L, Burkhardt T, Wisser J, Zimmermann R, Kurmanavicius J. Continuous independent quality control for fetal ultrasound biometry provided by the cumulative summation technique. *Ultrasound Obstet Gynecol* 2010;35(4):449-55. DOI: 10.1002/uog.7545.
171. Bardi F, Bakker M, Elvan-Taşpınar A, et al. Organ-specific learning curves of sonographers performing first-trimester anatomical screening and impact of score-based evaluation on ultrasound image quality. *PLoS One* 2023;18(2):e0279770. DOI: 10.1371/journal.pone.0279770.
172. Preeclampsia screening. Fetal Medicine Foundation. (<https://www.fetalmedicine.org/fmf-certification-2/preeclampsia-screening-1>).
173. Mathewlynn S, Starck LN, Yin Y, et al. Reference charts for first-trimester placental three-dimensional fractional moving blood volume derived using OxNNet. *Ultrasound Obstet Gynecol* 2026 DOI: 10.1002/uog.70161.
174. Mathewlynn S, Starck LN, Yin Y, et al. Reference charts for first-trimester placental volume derived using OxNNet. *Ultrasound Obstet Gynecol* 2025;66(3):337-346. DOI: 10.1002/uog.29300.
175. Page ES. Continuous inspection schemes. *Biometrika* 1954;41(1/2):100-115.
176. Lucas JM, Crosier RB. Fast initial response for CUSUM quality-control schemes: give your CUSUM a head start. *Technometrics* 2000;42(1):102-107.
177. Woodall WH, Montgomery DC. Some current directions in the theory and application of statistical process monitoring. *Journal of quality technology* 2014;46(1):78-94.
178. Chalouhi GE, Salomon LJ, Fontanges M, et al. Formative assessment based on an audit and feedback improves nuchal translucency ultrasound image quality. *J Ultrasound Med* 2013;32(9):1601-5. DOI: 10.7863/ultra.32.9.1601.
179. Torrent A, Manrique G, Gómez-Castelló T, et al. Sonologist's characteristics related to a higher quality in fetal nuchal translucency measured in primary antenatal care centers. *Prenat Diagn* 2019;39(10):934-939. DOI: 10.1002/pd.5512.
180. Cuckle H, Platt LD, Thornburg LL, et al. Nuchal Translucency Quality Review (NTQR) program: first one and half million results. *Ultrasound Obstet Gynecol* 2015;45(2):199-204. DOI: 10.1002/uog.13390.
181. Sabrià J, Barceló-Vidal C, Arigita M, Jimenez JM, Puerto B, Borrell A. The CUSUM test applied in prospective nuchal translucency quality review. *Ultrasound Obstet Gynecol* 2011;37(5):582-7. DOI: 10.1002/uog.8860.
182. Balsyte D, Schäffer L, Burkhardt T, Wisser J, Krafft A, Kurmanavicius J. Continuous independent quality control for fetal nuchal translucency measurements provided by the cumulative summation technique. *Ultraschall Med* 2011;32 Suppl 2:E141-6. DOI: 10.1055/s-0031-1281669.

183. Sahota DS, Leung WC, To WK, Chan WP, Lau TK, Leung TY. Quality assurance of nuchal translucency for prenatal fetal Down syndrome screening. *J Matern Fetal Neonatal Med* 2012;25(7):1039-43. DOI: 10.3109/14767058.2011.614658.
184. Weerasinghe S, Mirghani H, Revel A, Abu-Zidan FM. Cumulative sum (CUSUM) analysis in the assessment of trainee competence in fetal biometry measurement. *Ultrasound Obstet Gynecol* 2006;28(2):199-203. DOI: 10.1002/uog.2828.
185. Lee NMW, Chaemsaitong P, Poon LC. Prediction of preeclampsia in asymptomatic women. *Best Pract Res Clin Obstet Gynaecol* 2024;92:102436. DOI: 10.1016/j.bpobgyn.2023.102436.
186. Chaemsaitong P, Sahota DS, Poon LC. First trimester preeclampsia screening and prediction. *Am J Obstet Gynecol* 2022;226(2s):S1071-S1097.e2. DOI: 10.1016/j.ajog.2020.07.020.
187. Tan MY, Syngelaki A, Poon LC, et al. Screening for pre-eclampsia by maternal factors and biomarkers at 11-13 weeks' gestation. *Ultrasound Obstet Gynecol* 2018;52(2):186-195. DOI: 10.1002/uog.19112.
188. Tiruneh SA, Vu TTT, Moran LJ, et al. Externally validated prediction models for pre-eclampsia: systematic review and meta-analysis. *Ultrasound Obstet Gynecol* 2024;63(5):592-604. DOI: 10.1002/uog.27490.
189. Allotey J, Archer L, Coomaraswamy D, et al. Development and validation of prediction models for fetal growth restriction and birthweight: an individual participant data meta-analysis. *Health Technol Assess* 2024;28(47):1-119. DOI: 10.3310/dabw4814.
190. van Eekhout JCA, Becking EC, Scheffer PG, et al. First-Trimester Prediction Models Based on Maternal Characteristics for Adverse Pregnancy Outcomes: A Systematic Review and Meta-Analysis. *BJOG* 2025;132(3):243-265. DOI: 10.1111/1471-0528.17983.
191. Shah NS, Wang MC, Freaney PM, et al. Trends in Gestational Diabetes at First Live Birth by Race and Ethnicity in the US, 2011-2019. *Jama* 2021;326(7):660-669. DOI: 10.1001/jama.2021.7217.
192. Papastefanou I, Wright D, Syngelaki A, Souretis K, Chrysanthopoulou E, Nicolaides KH. Competing-risks model for prediction of small-for-gestational-age neonate from biophysical and biochemical markers at 11-13 weeks' gestation. *Ultrasound Obstet Gynecol* 2021;57(1):52-61. DOI: 10.1002/uog.23523.
193. Chaveeva P, Papastefanou I, Dagklis T, et al. External validation and comparison of Fetal Medicine Foundation competing-risks model for small-for-gestational-age neonate in the first trimester: multicenter cohort study. *Ultrasound Obstet Gynecol* 2025;65(6):729-737. DOI: 10.1002/uog.29219.
194. Wright D, Akolekar R, Syngelaki A, Poon LC, Nicolaides KH. A competing risks model in early screening for preeclampsia. *Fetal Diagn Ther* 2012;32(3):171-8. DOI: 10.1159/000338470.
195. O'Gorman N, Wright D, Poon LC, et al. Multicenter screening for pre-eclampsia by maternal factors and biomarkers at 11-13 weeks' gestation: comparison with NICE guidelines and ACOG recommendations. *Ultrasound Obstet Gynecol* 2017;49(6):756-760. DOI: 10.1002/uog.17455.

196. Chaemsaitong P, Pooh RK, Zheng M, et al. Prospective evaluation of screening performance of first-trimester prediction models for preterm preeclampsia in an Asian population. *Am J Obstet Gynecol* 2019;221(6):650.e1-650.e16. DOI: 10.1016/j.ajog.2019.09.041.
197. Tiruneh SA, Rolnik DL, Selvaratnam R, et al. External validation of the Fetal Medicine Foundation model for preterm pre-eclampsia prediction at 11-14 weeks in an Australian population. *Acta Obstet Gynecol Scand* 2025;104(9):1774-1782. DOI: 10.1111/aogs.70002.
198. Wright D, Tan MY, O'Gorman N, et al. Predictive performance of the competing risk model in screening for preeclampsia. *Am J Obstet Gynecol* 2019;220(2):199.e1-199.e13. DOI: 10.1016/j.ajog.2018.11.1087.
199. Papastefanou I, Chrelias C, Siristatidis C, Kappou D, Eleftheriades M, Kassanos D. Placental volume at 11 to 14 gestational weeks in pregnancies complicated with fetal growth restriction and preeclampsia. *Prenat Diagn* 2018;38(12):928-935. DOI: 10.1002/pd.5356.
200. González-González NL, González-Dávila E, González Marrero L, Padrón E, Conde JR, Plasencia W. Value of placental volume and vascular flow indices as predictors of intrauterine growth retardation. *Eur J Obstet Gynecol Reprod Biol* 2017;212:13-19. DOI: 10.1016/j.ejogrb.2017.03.005.
201. Wright D, Tan MY, O'Gorman N, Syngelaki A, Nicolaides KH. Serum PlGF compared with PAPP-A in first trimester screening for preterm pre-eclampsia: Adjusting for the effect of aspirin treatment. *BJOG* 2022;129(8):1308-1317. DOI: 10.1111/1471-0528.17096.
202. Liu HJ, Liu PC, Hua J, Zhao Y, Cao J. Placental weight and size in relation to fetal growth restriction: a case-control study. *J Matern Fetal Neonatal Med* 2021;34(9):1356-1360. DOI: 10.1080/14767058.2019.1636371.
203. Vachon-Marceau C, Demers S, Markey S, et al. First-trimester placental thickness and the risk of preeclampsia or SGA. *Placenta* 2017;57:123-128. DOI: 10.1016/j.placenta.2017.06.016.
204. Stevenson GN, Noble JA, Welsh AW, Impey L, Collins SL. Automated Visualization and Quantification of Spiral Artery Blood Flow Entering the First-Trimester Placenta, Using 3-D Power Doppler Ultrasound. *Ultrasound Med Biol* 2018;44(3):522-531. DOI: 10.1016/j.ultrasmedbio.2017.11.018.
205. Lorensen WE, Cline HE. Marching cubes: A high resolution 3D surface construction algorithm. *Seminal graphics: pioneering efforts that shaped the field* 1998:347-353.
206. Van der Walt S, Schönberger JL, Nunez-Iglesias J, et al. scikit-image: image processing in Python. *PeerJ* 2014;2:e453.
207. Lee T-C, Kashyap RL, Chu C-N. Building skeleton models via 3-D medial surface axis thinning algorithms. *CVGIP: graphical models and image processing* 1994;56(6):462-478.
208. Salavati N, Sovio U, Mayo RP, Charnock-Jones DS, Smith GC. The relationship between human placental morphometry and ultrasonic measurements of utero-placental blood flow and fetal growth. *Placenta* 2016;38:41-8. DOI: 10.1016/j.placenta.2015.12.003.
209. Suri S, Muttukrishna S, Jauniaux E. 2D-Ultrasound and endocrinologic evaluation of placentation in early pregnancy and its relationship to fetal birthweight in normal

- pregnancies and pre-eclampsia. *Placenta* 2013;34(9):745-50. DOI: 10.1016/j.placenta.2013.05.003.
210. Mayhew TM, Manwani R, Ohadike C, Wijesekara J, Baker PN. The placenta in pre-eclampsia and intrauterine growth restriction: studies on exchange surface areas, diffusion distances and villous membrane diffusive conductances. *Placenta* 2007;28(2-3):233-8. DOI: 10.1016/j.placenta.2006.02.011.
 211. Mayhew TM, Ohadike C, Baker PN, Crocker IP, Mitchell C, Ong SS. Stereological investigation of placental morphology in pregnancies complicated by pre-eclampsia with and without intrauterine growth restriction. *Placenta* 2003;24(2-3):219-26. DOI: 10.1053/plac.2002.0900.
 212. Ong SS, Baker PN, Mayhew TM, Dunn WR. Remodeling of myometrial radial arteries in preeclampsia. *Am J Obstet Gynecol* 2005;192(2):572-9. DOI: 10.1016/j.ajog.2004.08.015.
 213. Allerkamp HH, Clark AR, Lee TC, Morgan TK, Burton GJ, James JL. Something old, something new: digital quantification of uterine vascular remodelling and trophoblast plugging in historical collections provides new insight into adaptation of the utero-placental circulation. *Hum Reprod* 2021;36(3):571-586. DOI: 10.1093/humrep/deaa303.
 214. Zamir M, Nelson DM, Ginosar Y. Hemodynamic consequences of incomplete uterine spiral artery transformation in human pregnancy, with implications for placental dysfunction and preeclampsia. *J Appl Physiol* (1985) 2021;130(2):457-465. DOI: 10.1152/jappphysiol.00504.2020.
 215. Roth CJ, Haeussner E, Ruebelmann T, et al. Dynamic modeling of uteroplacental blood flow in IUGR indicates vortices and elevated pressure in the intervillous space - a pilot study. *Sci Rep* 2017;7:40771. DOI: 10.1038/srep40771.
 216. Esplin MS, Faurett MB, Fraser A, et al. Paternal and maternal components of the predisposition to preeclampsia. *N Engl J Med* 2001;344(12):867-72. DOI: 10.1056/nejm200103223441201.
 217. Melamed B, Aviram A, Barg M, Mei-Dan E. The smaller firstborn: exploring the association of parity and fetal growth. *Arch Gynecol Obstet* 2024;310(1):93-102. DOI: 10.1007/s00404-023-07249-5.
 218. Johnson JD, Louis JM. Does race or ethnicity play a role in the origin, pathophysiology, and outcomes of preeclampsia? An expert review of the literature. *Am J Obstet Gynecol* 2022;226(2s):S876-s885. DOI: 10.1016/j.ajog.2020.07.038.
 219. Duckitt K, Harrington D. Risk factors for pre-eclampsia at antenatal booking: systematic review of controlled studies. *Bmj* 2005;330(7491):565. DOI: 10.1136/bmj.38380.674340.E0.
 220. Henderson JT, Vesco KK, Senger CA, Thomas RG, Redmond N. Aspirin Use to Prevent Preeclampsia and Related Morbidity and Mortality: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *Jama* 2021;326(12):1192-1206. DOI: 10.1001/jama.2021.8551.
 221. Pajkrt E, Mol BW, Boer K, Drogtróp AP, Bossuyt PM, Bilardo CM. Intra- and interoperator repeatability of the nuchal translucency measurement. *Ultrasound Obstet Gynecol* 2000;15(4):297-301. DOI: 10.1046/j.1469-0705.2000.00088.x.

222. Axell RG, Gillett A, Pasupathy D, et al. Accuracy of nuchal translucency measurement depends on equipment used and its calibration. *Ultrasound Obstet Gynecol* 2014;44(1):31-7. DOI: 10.1002/uog.13306.
223. Prins LI, Smith E, Naaktgeboren CM, et al. Inter- and intra-observer reliability in first-trimester uterine artery Doppler measurements. *Eur J Obstet Gynecol Reprod Biol* 2025;311:114058. DOI: 10.1016/j.ejogrb.2025.114058.
224. Ferreira AE, Mauad Filho F, Abreu PS, Mauad FM, Araujo Júnior E, Martins WP. Reproducibility of first- and second-trimester uterine artery pulsatility index measured by transvaginal and transabdominal ultrasound. *Ultrasound Obstet Gynecol* 2015;46(5):546-52. DOI: 10.1002/uog.14762.
225. Marchi L, Zwertbroek E, Snelder J, Kloosterman M, Bilardo CM. Intra- and inter-observer reproducibility and generalizability of first trimester uterine artery pulsatility index by transabdominal and transvaginal ultrasound. *Prenat Diagn* 2016;36(13):1261-1269. DOI: 10.1002/pd.4970.
226. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 2002;21(9):1331-5. DOI: 10.1002/sim.1108.
227. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86(2):420-8. DOI: 10.1037//0033-2909.86.2.420.
228. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological methods* 1996;1(1):30.
229. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine* 2016;15(2):155-163.
230. Cabezas López E, Martínez-Payo C, Engels Calvo V, San Frutos Llorente L, Pérez-Medina T. Reproducibility of first trimester three-dimensional placental measurements. *Eur J Obstet Gynecol Reprod Biol* 2016;201:156-60. DOI: 10.1016/j.ejogrb.2016.03.034.
231. Nardoza LM, Nowak PM, Araujo Júnior E, et al. Evaluation of placental volume at 7-10+6 weeks of pregnancy by 3D-sonography. *Placenta* 2009;30(7):585-9. DOI: 10.1016/j.placenta.2009.04.008.
232. Huster KM, Haas K, Schoenborn J, McVean D, Odibo AO. Reproducibility of placental volume and vasculature indices obtained by 3-dimensional power Doppler sonography. *J Ultrasound Med* 2010;29(6):911-6. DOI: 10.7863/jum.2010.29.6.911.

[10] APPENDICES

Appendix 1: Cumulative sum (CUSUM) learning curve analyses for each sonographer performance with respect to subgroup cohort 1 for the entire duration of the OxPLUS study

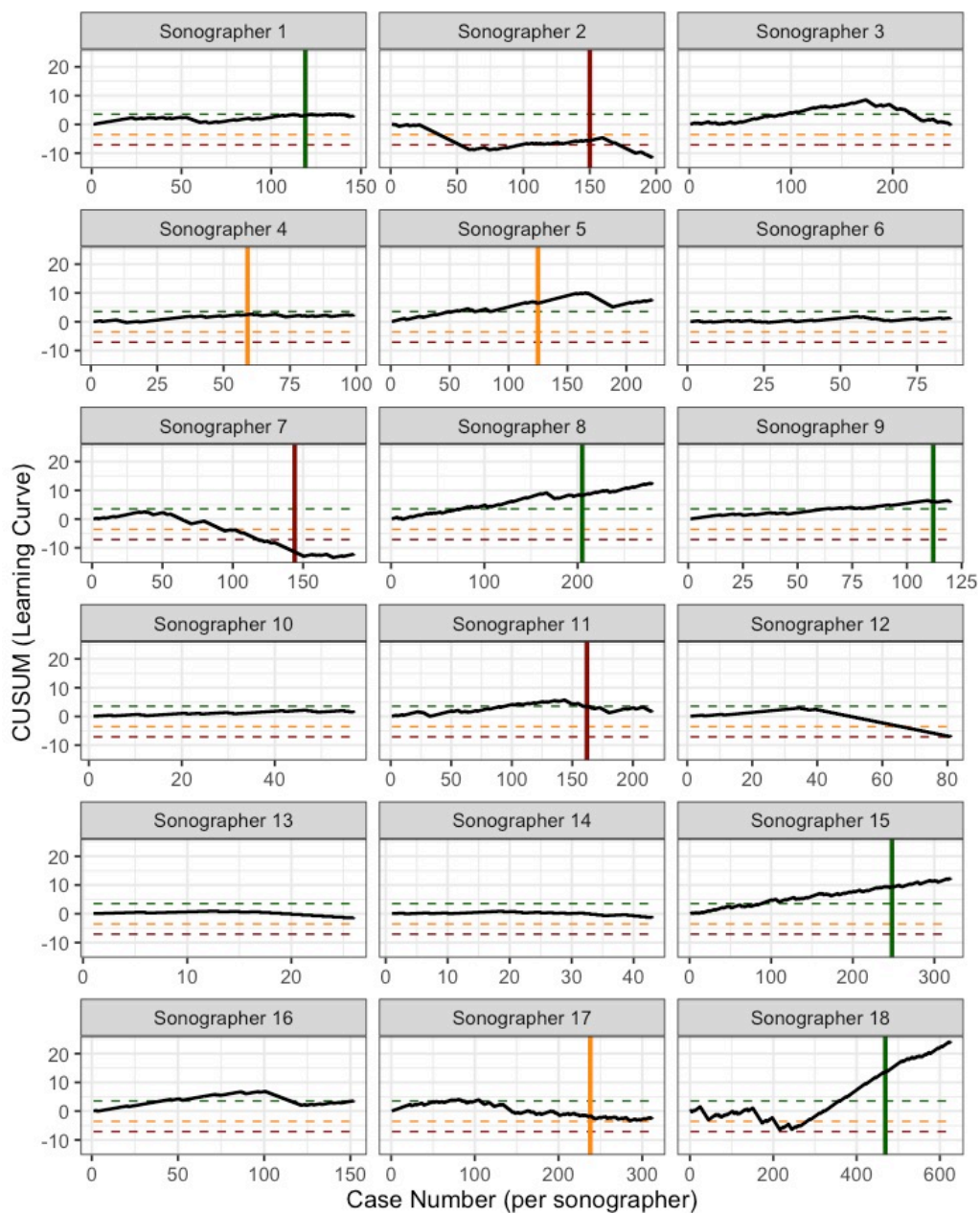
Incremental increases in the CUSUM curve represent improved performance.

Green dashed horizontal line (top): Reassurance threshold

Amber dashed horizontal line (middle): Warning threshold

Red dashed horizontal line (bottom): Alarm threshold

Solid vertical lines: The time at which the feedback was delivered using the OxNNet feedback toolkit; colour coded to represent pass (green), advisory only (amber), or failure in any domain (red).



Appendix 2: Cumulative sum (CUSUM) learning curve analyses for each sonographer performance with respect to subgroup cohort 2 for the entire duration of the OxPLUS study

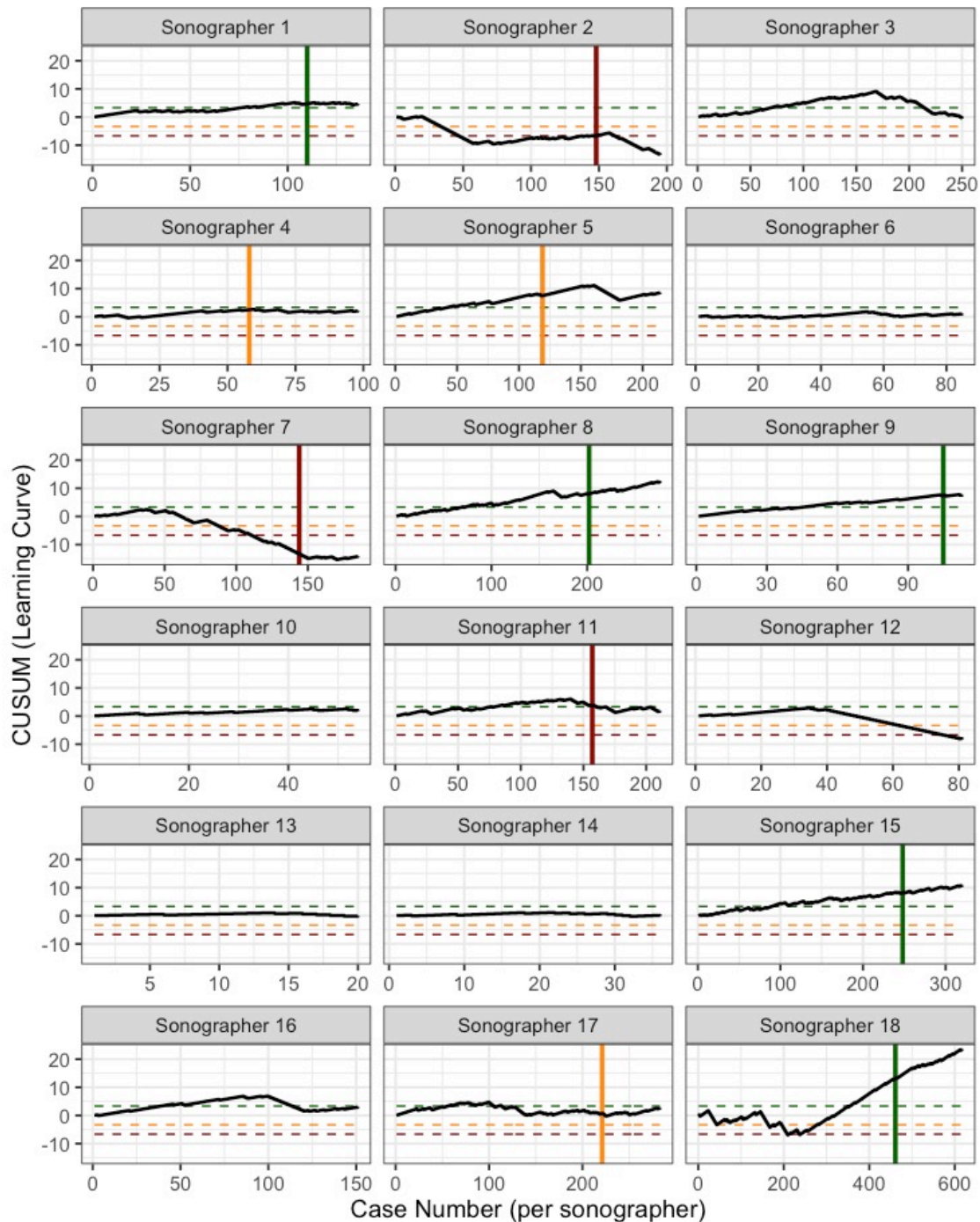
Incremental increases in the CUSUM curve represent improved performance.

Green dashed horizontal line (top): Reassurance threshold

Amber dashed horizontal line (middle): Warning threshold

Red dashed horizontal line (bottom): Alarm threshold

Solid vertical lines: The time at which the feedback was delivered using the OxNNet feedback toolkit; colour coded to represent pass (green), advisory only (amber), or failure in any domain (red).



Appendix 3: Candidate variables plotted against log relative hazard (with 95% confidence intervals) for each outcome variable, exploring linear modelling and restricted cubic splines with 2 or 3 degrees of freedom.

Appendix 3a: Maternal age

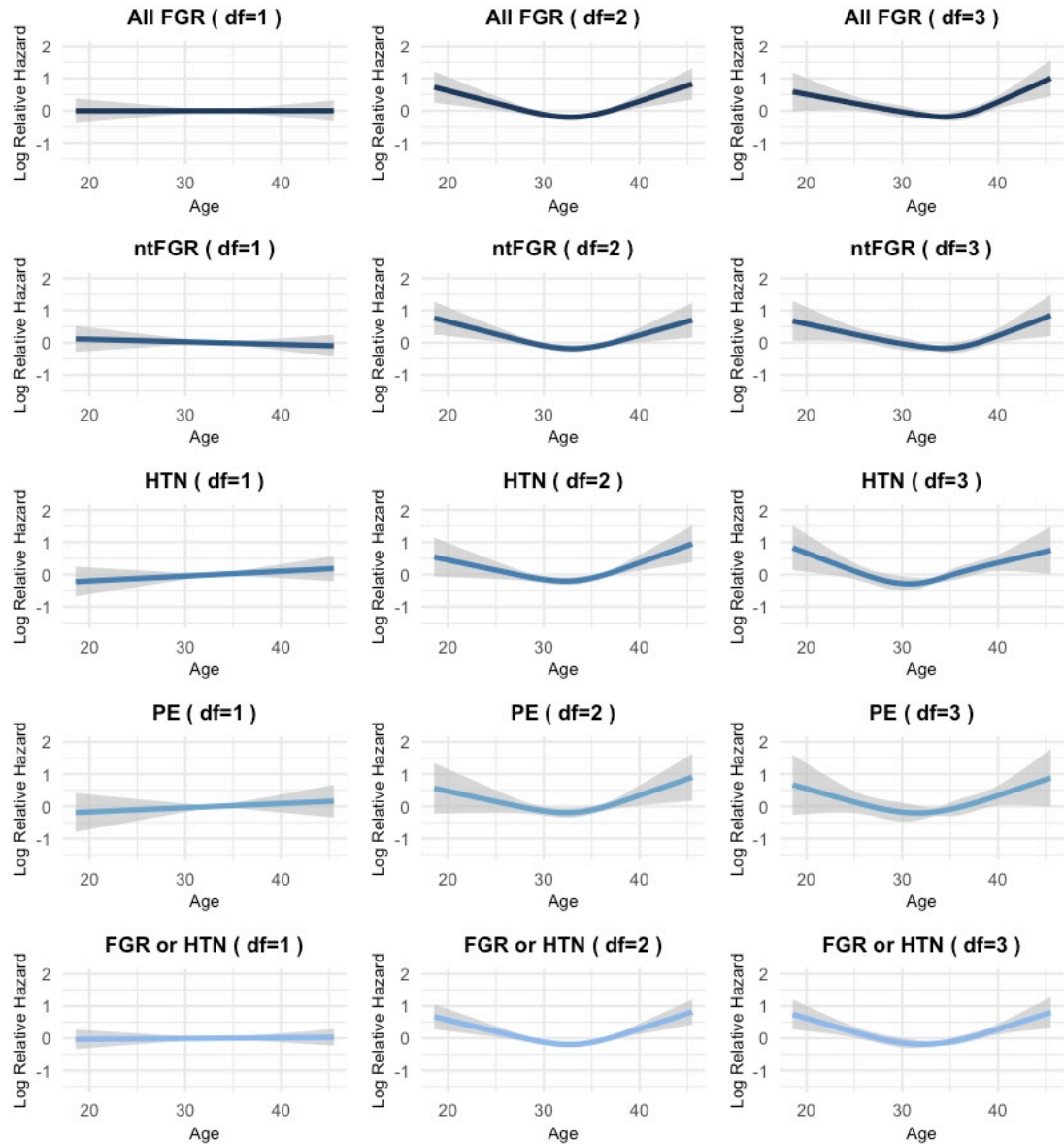
FGR: Fetal growth restriction

ntFGR: Normotensive fetal growth restriction

HTN: Any gestational hypertension, including PIH and PE

PE: Pre-eclampsia

FGR or HTN: A composite outcome of any FGR or HTN



Appendix 3b: Last birthweight percentile

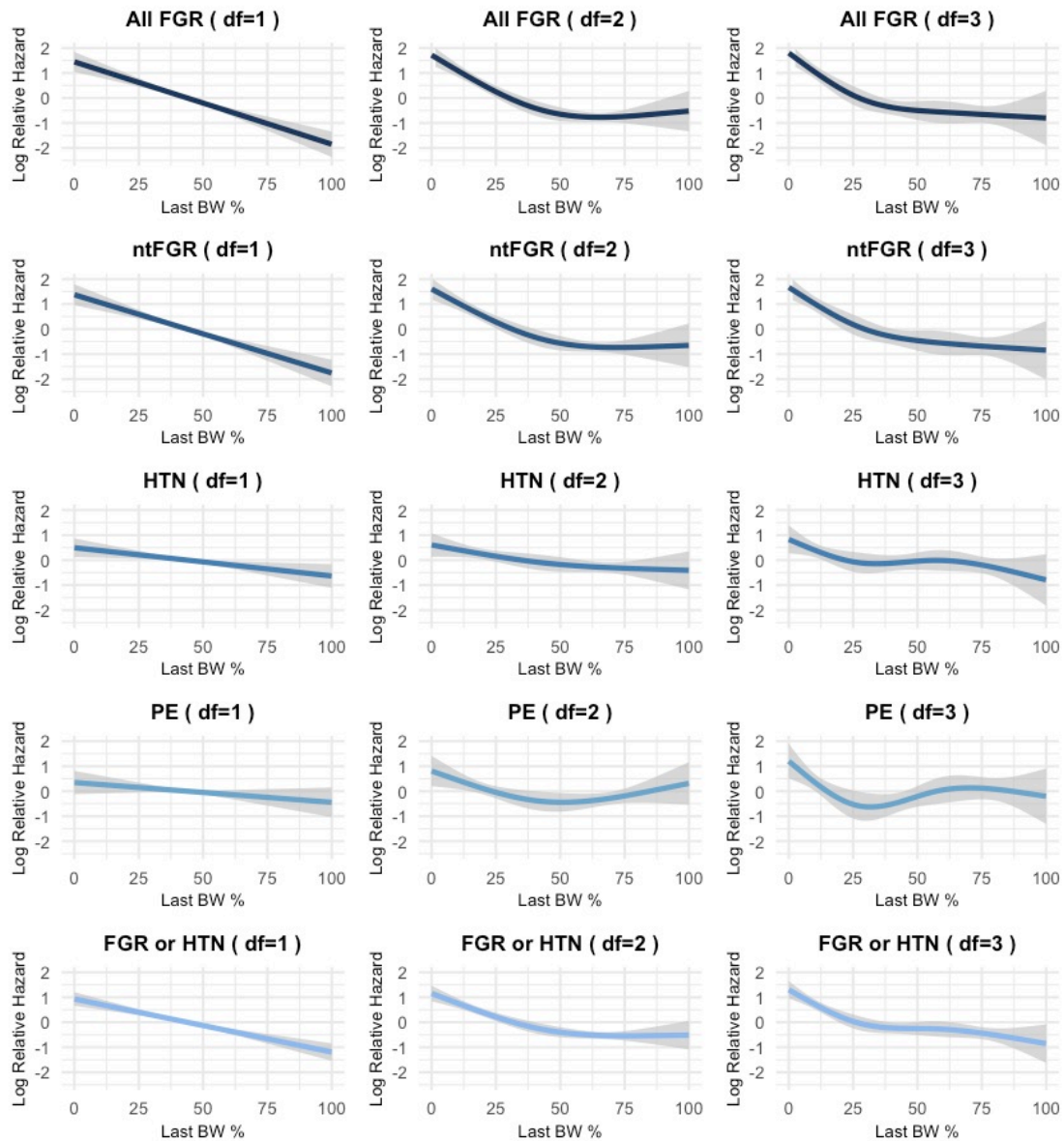
FGR: Fetal growth restriction

ntFGR: Normotensive fetal growth restriction

HTN: Any gestational hypertension, including PIH and PE

PE: Pre-eclampsia

FGR or HTN: A composite outcome of any FGR or HTN



Appendix 3c: $\log_{10}$ serum PAPP-A MoMs

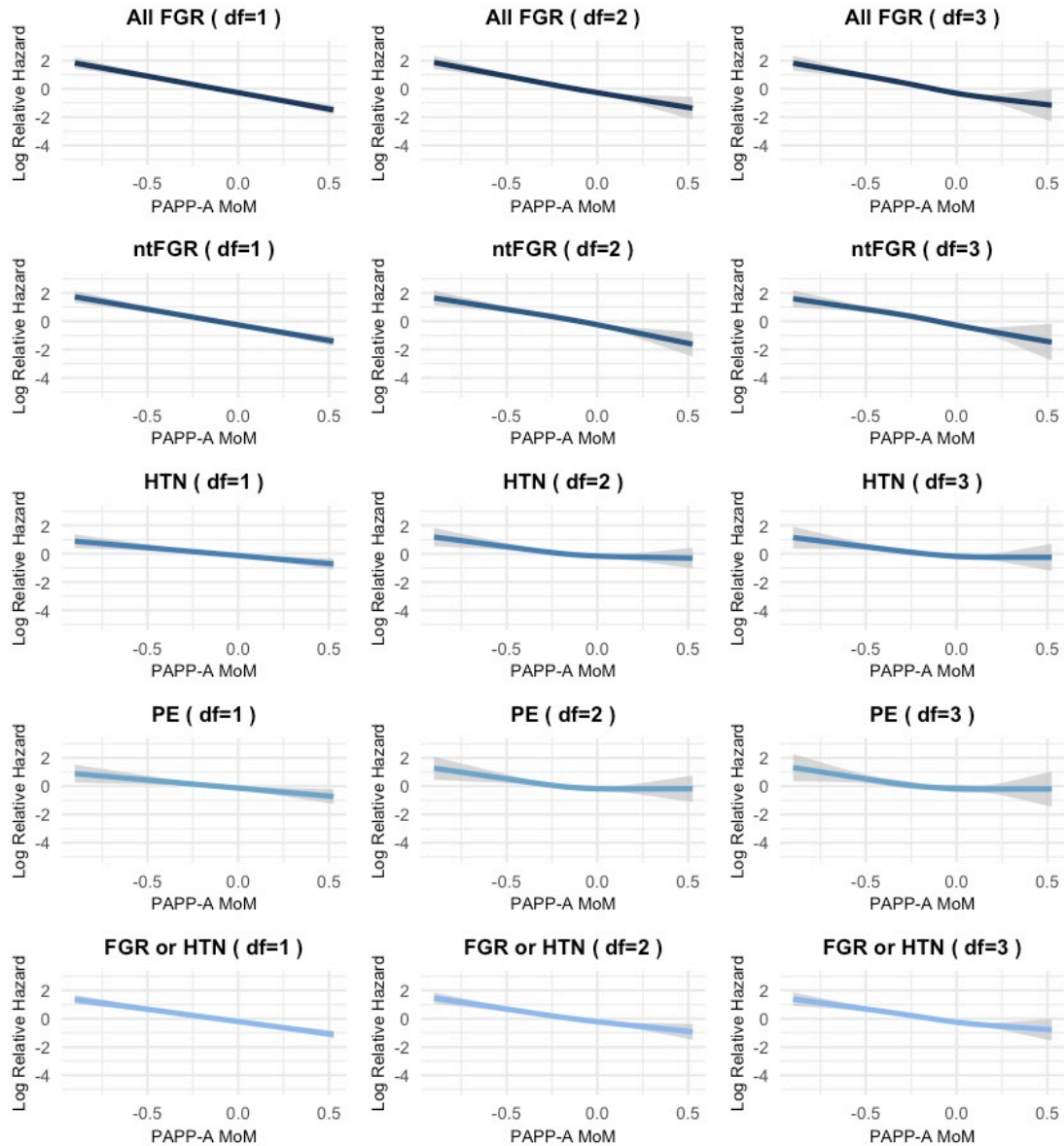
FGR: Fetal growth restriction

ntFGR: Normotensive fetal growth restriction

HTN: Any gestational hypertension, including PIH and PE

PE: Pre-eclampsia

FGR or HTN: A composite outcome of any FGR or HTN



Appendix 3d: $\log_{10}$ serum PLGF MoMs

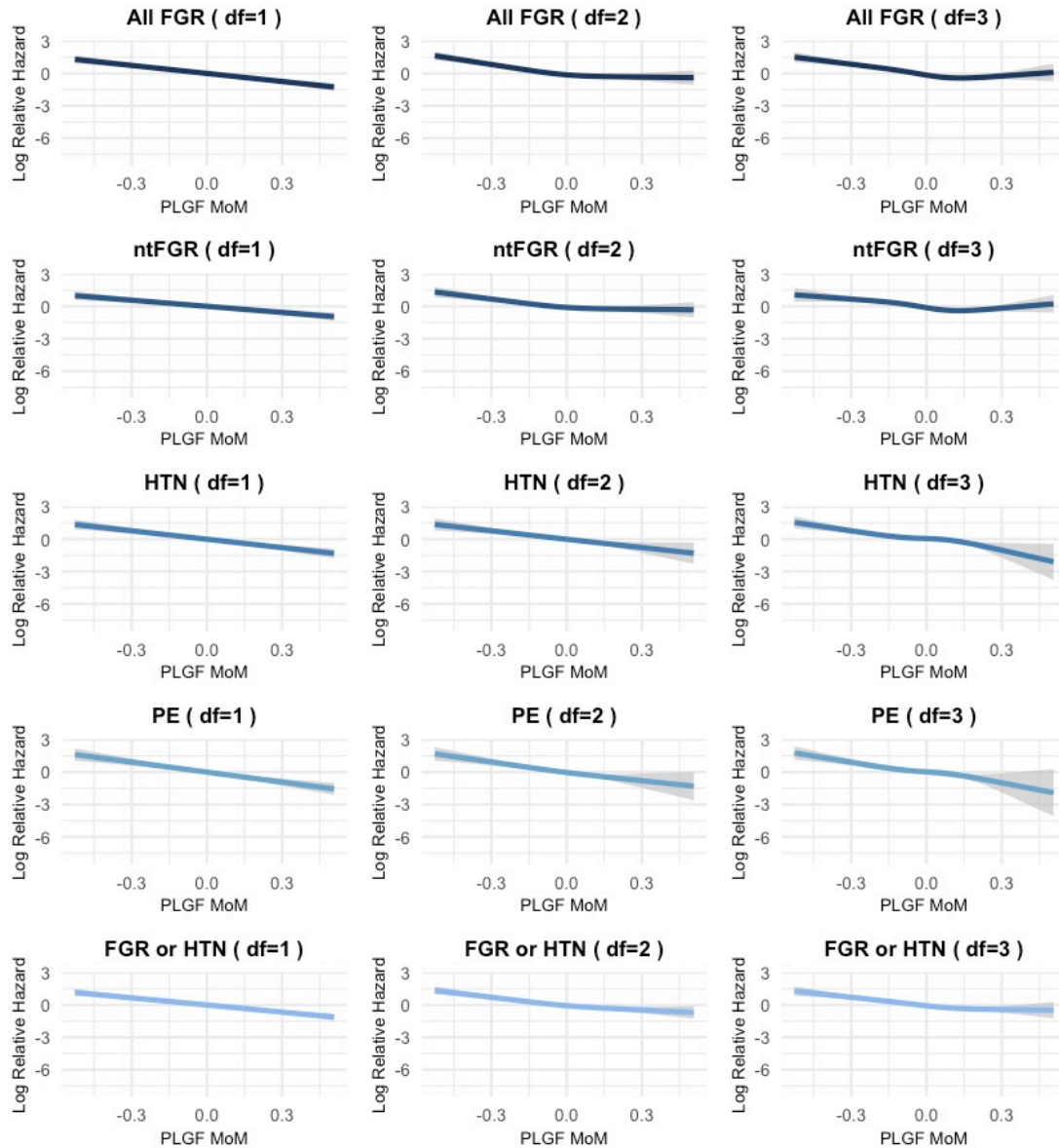
FGR: Fetal growth restriction

ntFGR: Normotensive fetal growth restriction

HTN: Any gestational hypertension, including PIH and PE

PE: Pre-eclampsia

FGR or HTN: A composite outcome of any FGR or HTN



Appendix 3e: $\log_{10}$ mean arterial pressure MoMs

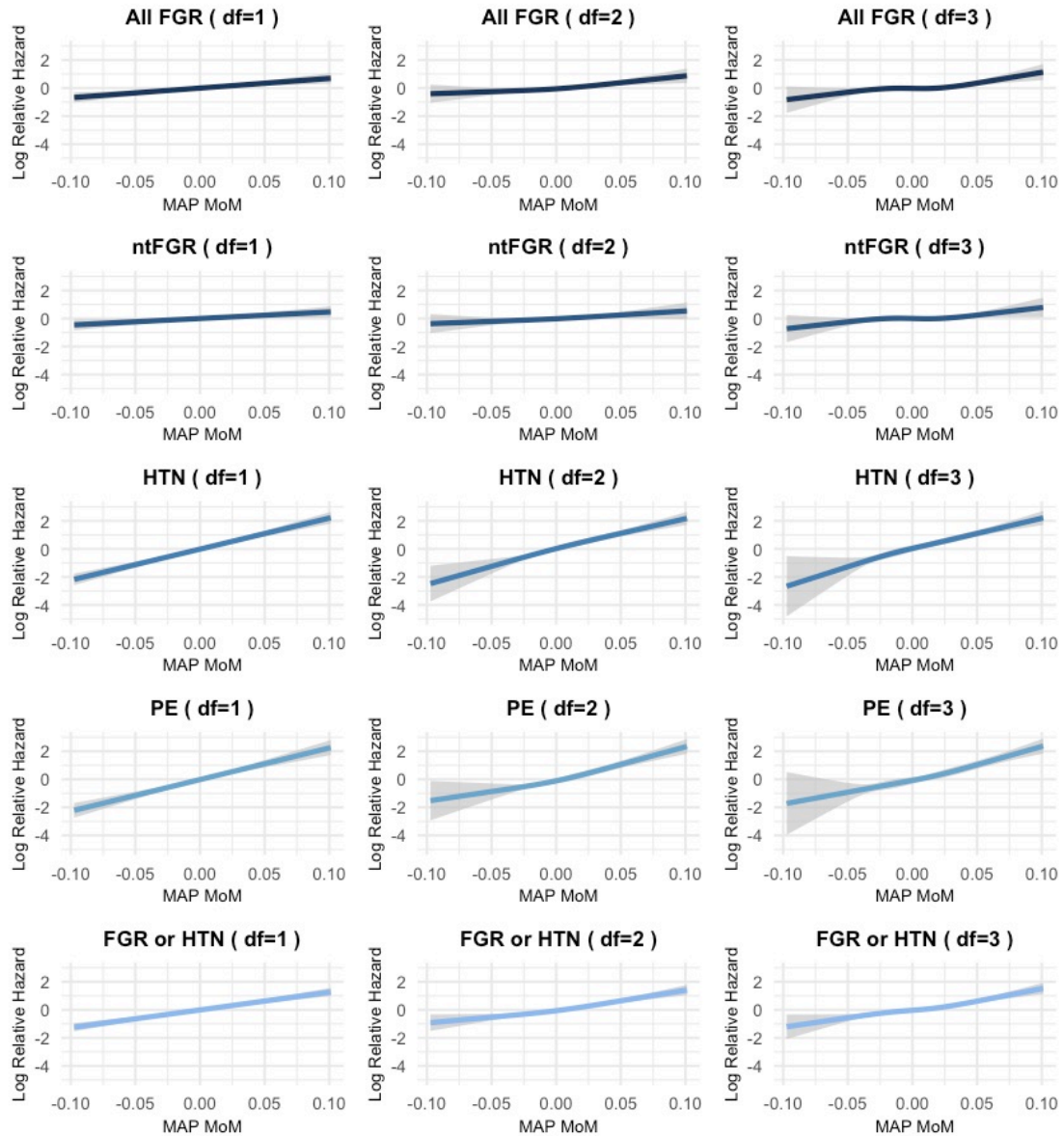
FGR: Fetal growth restriction

ntFGR: Normotensive fetal growth restriction

HTN: Any gestational hypertension, including PIH and PE

PE: Pre-eclampsia

FGR or HTN: A composite outcome of any FGR or HTN



Appendix 3f: $\log_{10}$ uterine artery PI MoMs

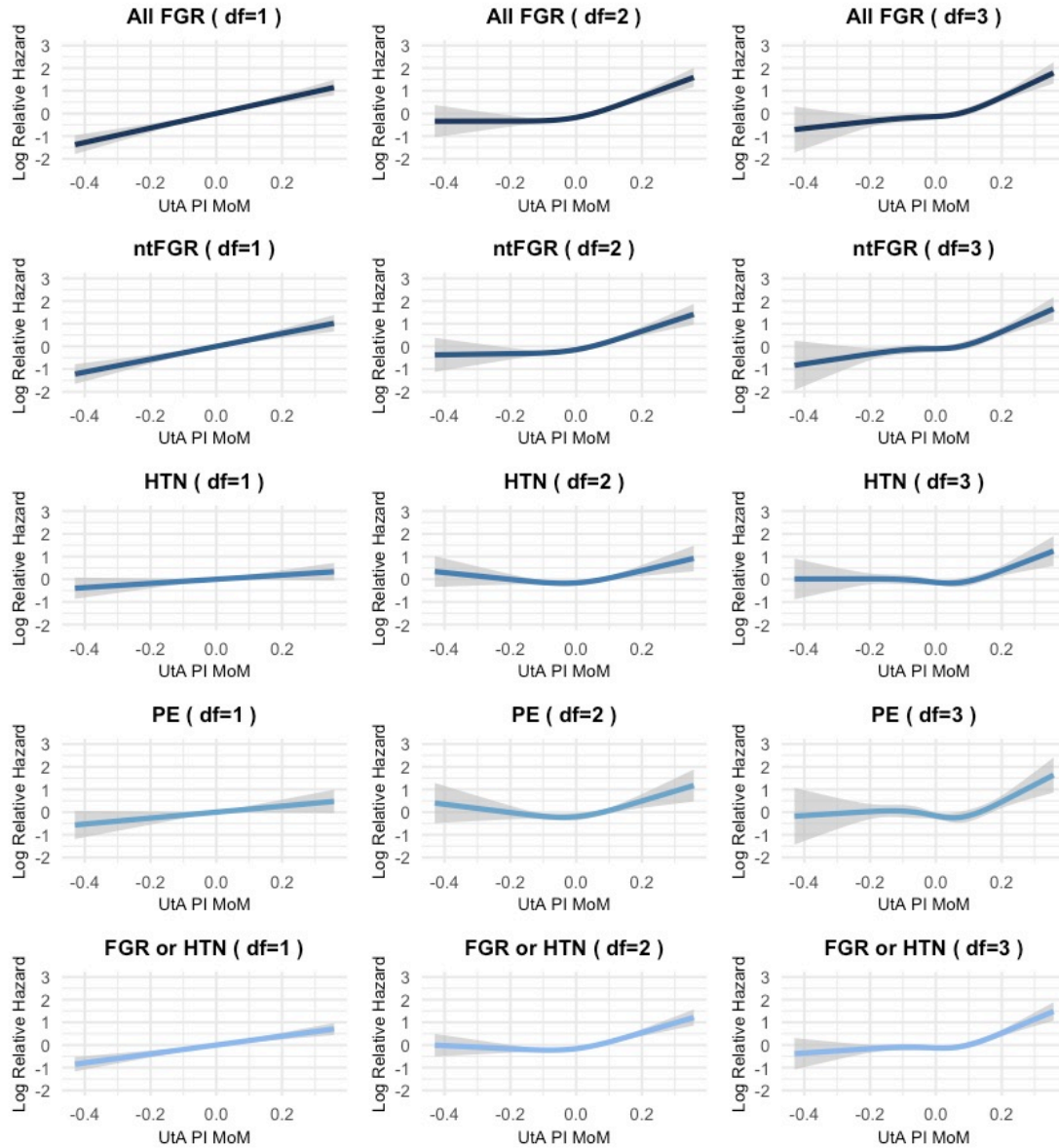
FGR: Fetal growth restriction

ntFGR: Normotensive fetal growth restriction

HTN: Any gestational hypertension, including PIH and PE

PE: Pre-eclampsia

FGR or HTN: A composite outcome of any FGR or HTN



Appendix 3g: svFMBV5 z-score

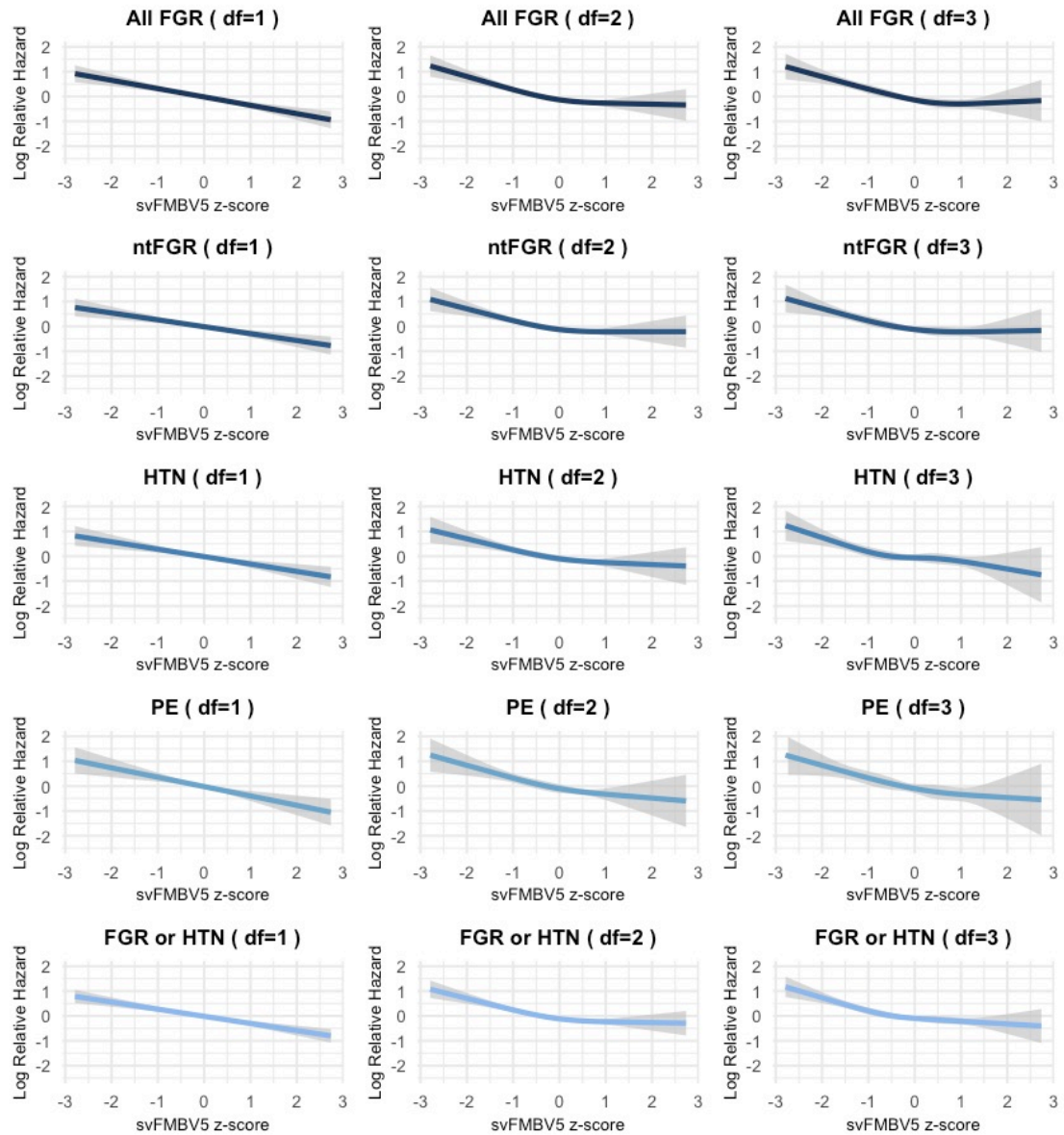
FGR: Fetal growth restriction

ntFGR: Normotensive fetal growth restriction

HTN: Any gestational hypertension, including PIH and PE

PE: Pre-eclampsia

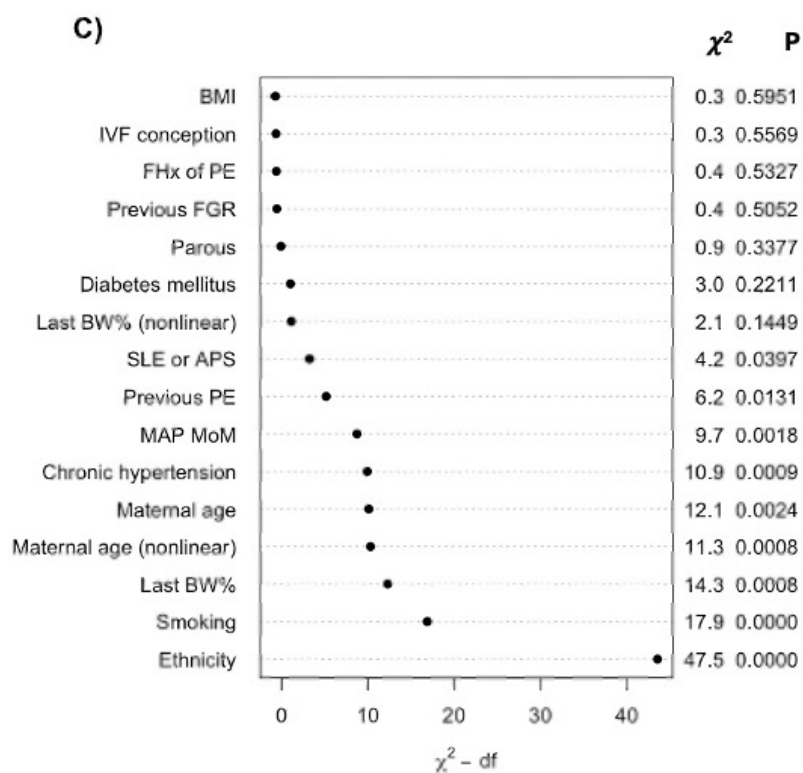
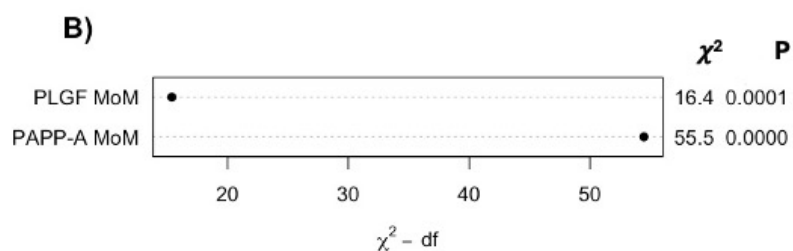
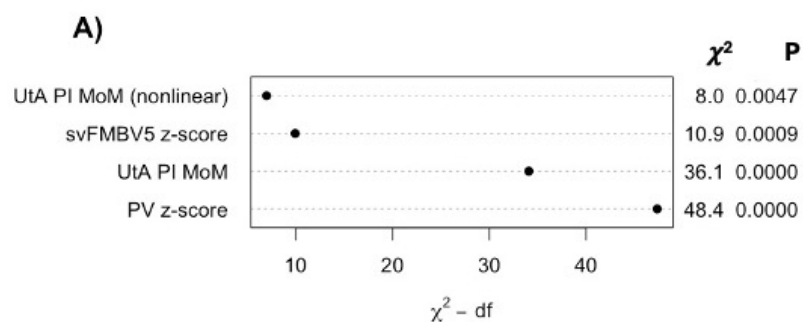
FGR or HTN: A composite outcome of any FGR or HTN



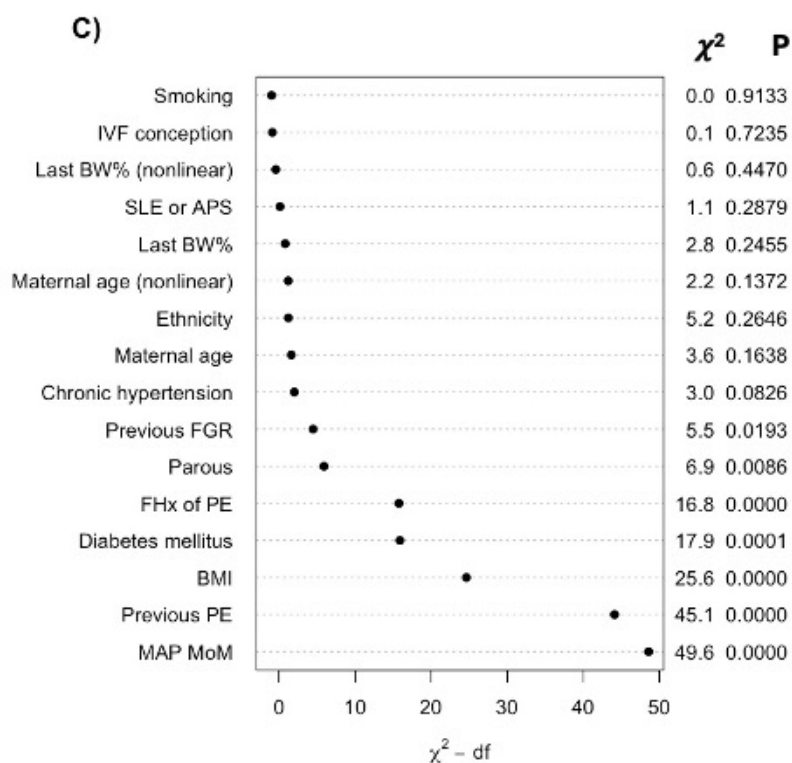
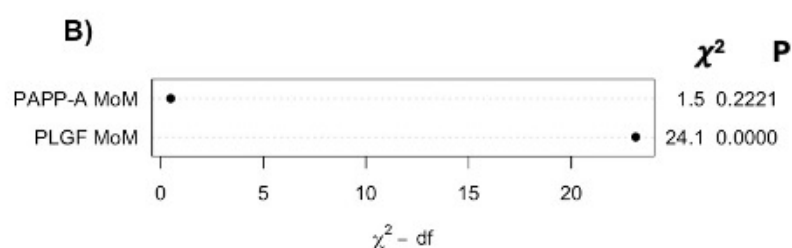
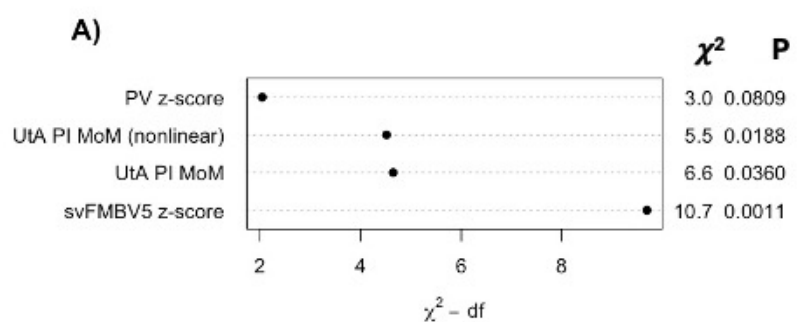
Appendix 4: AIC values for univariable prediction of fetal growth restriction and pre-eclampsia exploring restricted cubic splines with varying degrees of freedom

Outcome	Predictor	AIC value for variable degrees of freedom		
		1DF	2DF	3DF
FGR	Maternal age	3828	3816	3817
	svFMBV5 z-score	3800	3798	3799
	Last BW percentile	1369	1360	1361
	log ₁₀ MAP MoM	3815	3816	3816
	log ₁₀ PAPP-A MoM	3749	3751	3752
	log ₁₀ PlGF MoM	3786	3781	3780
	log ₁₀ UtA PI MoM	3784	3777	3777
	Placental volume z-score	3755	3757	3758
PE	Maternal age	1594	1590	1592
	svFMBV5 z-score	1579	1580	1582
	Last BW percentile	620	617	615
	log ₁₀ MAP MoM	1532	1533	1535
	log ₁₀ PAPP-A MoM	1586	1587	1589
	log ₁₀ PlGF MoM	1566	1568	1569
	log ₁₀ UtA PI MoM	1591	1587	1585
	Placental volume z-score	1588	1588	1586

Appendix 5: ANOVA plots generated during backwards selection process for prediction of fetal growth restriction, including A) sonographic markers, B) serum biomarkers, and C) mean arterial blood pressure and factors from the maternal medical history



Appendix 6: ANOVA plots generated during backwards selection process for prediction of pre-eclampsia, including A) sonographic markers, B) serum biomarkers, and C) mean arterial blood pressure and factors from the maternal medical history



Appendix 7: Sensitivity, specificity, negative predictive value and positive predictive value for the performance of each model in each subgroup on external validation

Subgroup 1: Excluding term FGR and term PE

Model	10% FPR			
	Sens	Spec	PPV	NPV
FGR1	0.29 (0.14–0.48)	0.90 (0.89–0.91)	0.03 (0.02–0.06)	0.99 (0.99–0.99)
FGR2	0.23 (0.10–0.41)	0.90 (0.89–0.91)	0.03 (0.01–0.05)	0.99 (0.99–0.99)
FGR3	0.26 (0.12–0.45)	0.90 (0.89–0.91)	0.03 (0.01–0.06)	0.99 (0.99–0.99)
FGR4	0.26 (0.12–0.45)	0.90 (0.89–0.91)	0.03 (0.01–0.06)	0.99 (0.99–0.99)
PE1	0.46 (0.27–0.67)	0.90 (0.89–0.91)	0.04 (0.02–0.07)	0.99 (0.99–1.00)
PE2	0.23 (0.09–0.44)	0.90 (0.89–0.91)	0.02 (0.01–0.05)	0.99 (0.99–0.99)
PE3	0.31 (0.14–0.52)	0.90 (0.89–0.91)	0.03 (0.01–0.06)	0.99 (0.99–1.00)
PE4	0.35 (0.17–0.56)	0.90 (0.89–0.91)	0.03 (0.02–0.06)	0.99 (0.99–1.00)

Model	30% FPR			
	Sens	Spec	PPV	NPV
FGR1	0.65 (0.45–0.81)	0.70 (0.68–0.72)	0.02 (0.02–0.04)	0.99 (0.99–1.00)
FGR2	0.58 (0.39–0.75)	0.70 (0.68–0.72)	0.02 (0.01–0.03)	0.99 (0.99–1.00)
FGR3	0.45 (0.27–0.64)	0.70 (0.68–0.72)	0.02 (0.01–0.03)	0.99 (0.99–0.99)
FGR4	0.71 (0.52–0.86)	0.70 (0.68–0.72)	0.03 (0.02–0.04)	1.00 (0.99–1.00)
PE1	0.85 (0.65–0.96)	0.70 (0.68–0.72)	0.03 (0.02–0.04)	1.00 (0.99–1.00)
PE2	0.62 (0.41–0.80)	0.70 (0.68–0.72)	0.02 (0.01–0.03)	0.99 (0.99–1.00)
PE3	0.62 (0.41–0.80)	0.68 (0.66–0.70)	0.02 (0.01–0.03)	0.99 (0.99–1.00)
PE4	0.65 (0.44–0.83)	0.70 (0.68–0.72)	0.02 (0.01–0.03)	1.00 (0.99–1.00)

Model	50% FPR			
	Sens	Spec	PPV	NPV
FGR1	0.84 (0.66–0.95)	0.50 (0.48–0.52)	0.02 (0.01–0.03)	1.00 (0.99–1.00)
FGR2	0.74 (0.55–0.88)	0.50 (0.48–0.52)	0.02 (0.01–0.03)	0.99 (0.99–1.00)
FGR3	0.65 (0.45–0.81)	0.50 (0.48–0.52)	0.01 (0.01–0.02)	0.99 (0.99–1.00)
FGR4	0.74 (0.55–0.88)	0.50 (0.48–0.52)	0.02 (0.01–0.03)	0.99 (0.99–1.00)
PE1	0.92 (0.75–0.99)	0.50 (0.48–0.52)	0.02 (0.01–0.03)	1.00 (0.99–1.00)
PE2	0.73 (0.52–0.88)	0.50 (0.48–0.52)	0.01 (0.01–0.02)	0.99 (0.99–1.00)
PE3	0.69 (0.48–0.86)	0.48 (0.46–0.50)	0.01 (0.01–0.02)	0.99 (0.99–1.00)
PE4	0.81 (0.61–0.93)	0.50 (0.48–0.52)	0.02 (0.01–0.02)	1.00 (0.99–1.00)

Subgroup 2: Pre-eclampsia cases excluded

Model	10% FPR			
	Sens	Spec	PPV	NPV
FGR1	0.33 (0.24–0.44)	0.90 (0.88–0.91)	0.11 (0.07–0.15)	0.97 (0.97–0.98)
FGR2	0.24 (0.15–0.34)	0.90 (0.89–0.91)	0.08 (0.05–0.12)	0.97 (0.96–0.98)
FGR3	0.40 (0.30–0.50)	0.89 (0.88–0.90)	0.12 (0.08–0.16)	0.98 (0.97–0.98)
FGR4	0.39 (0.29–0.49)	0.89 (0.88–0.90)	0.11 (0.08–0.15)	0.98 (0.97–0.98)

Model	30% FPR			
	Sens	Spec	PPV	NPV
FGR1	0.61 (0.51–0.71)	0.70 (0.68–0.72)	0.07 (0.05–0.09)	0.98 (0.97–0.99)
FGR2	0.48 (0.38–0.59)	0.70 (0.68–0.72)	0.06 (0.04–0.07)	0.97 (0.97–0.98)
FGR3	0.58 (0.47–0.68)	0.70 (0.68–0.72)	0.07 (0.05–0.09)	0.98 (0.97–0.98)
FGR4	0.66 (0.55–0.75)	0.70 (0.68–0.72)	0.07 (0.06–0.09)	0.98 (0.97–0.99)

Model	50% FPR			
	Sens	Spec	PPV	NPV
FGR1	0.87 (0.79–0.93)	0.49 (0.47–0.51)	0.06 (0.05–0.07)	0.99 (0.98–1.00)
FGR2	0.62 (0.52–0.72)	0.50 (0.48–0.52)	0.04 (0.03–0.06)	0.97 (0.96–0.98)
FGR3	0.73 (0.63–0.82)	0.51 (0.49–0.52)	0.05 (0.04–0.06)	0.98 (0.97–0.99)
FGR4	0.74 (0.64–0.83)	0.51 (0.49–0.53)	0.05 (0.04–0.07)	0.98 (0.97–0.99)

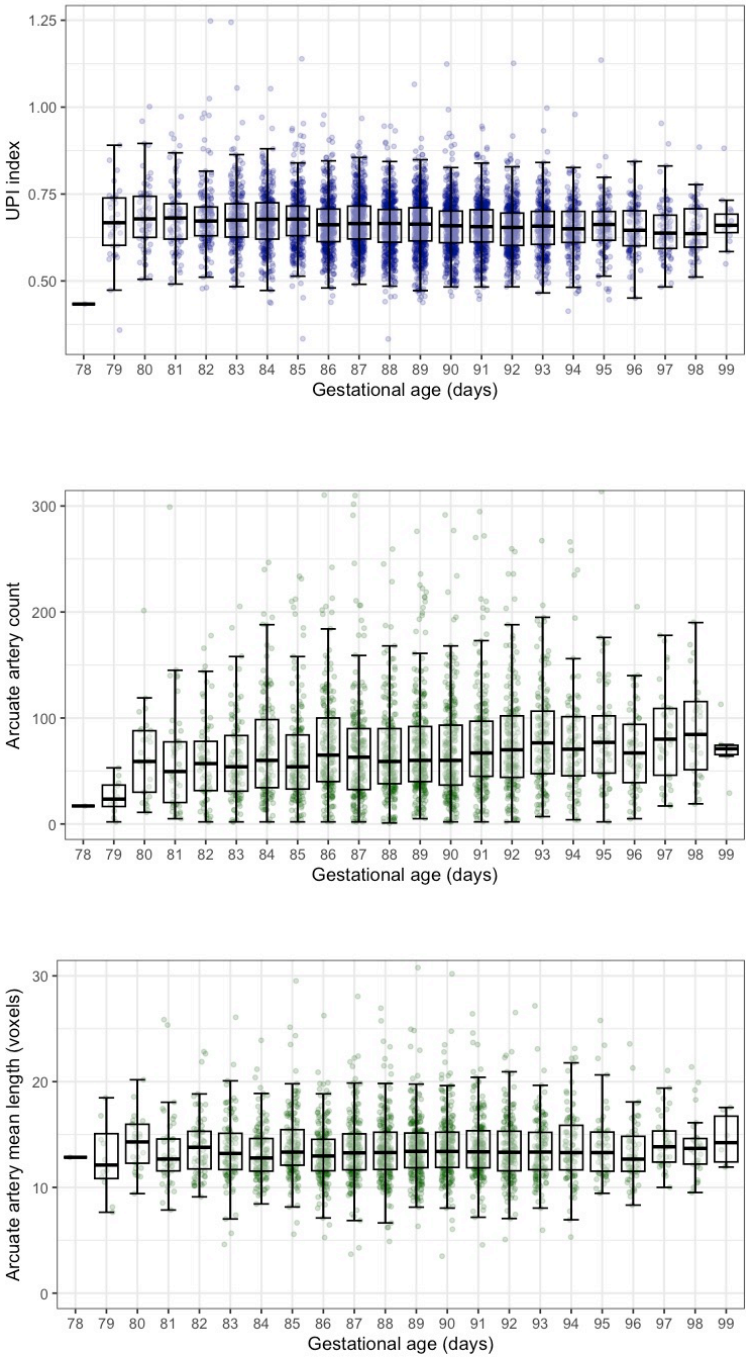
Subgroup 3: Pre-existing and gestational diabetes excluded

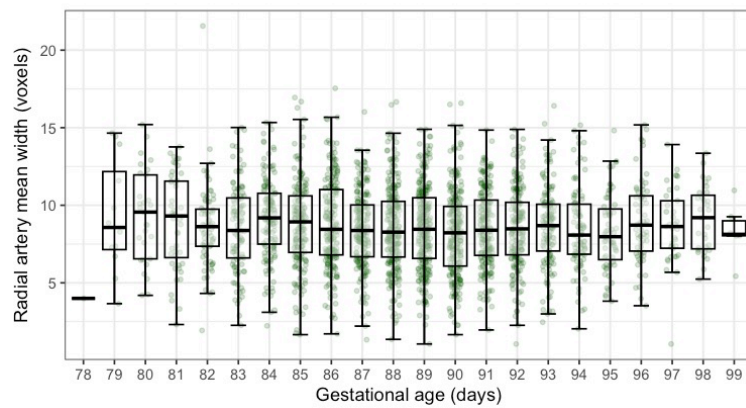
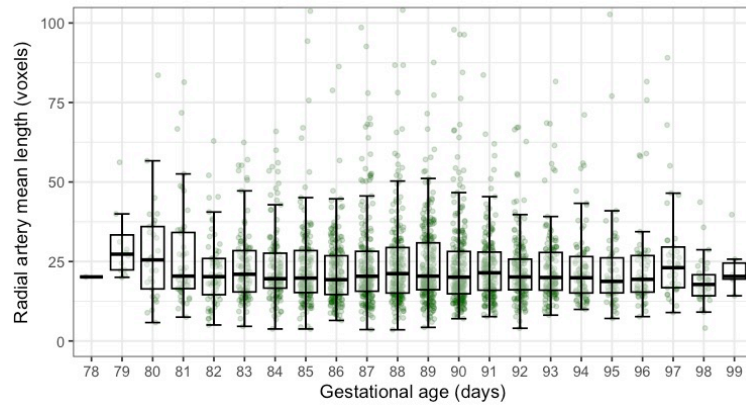
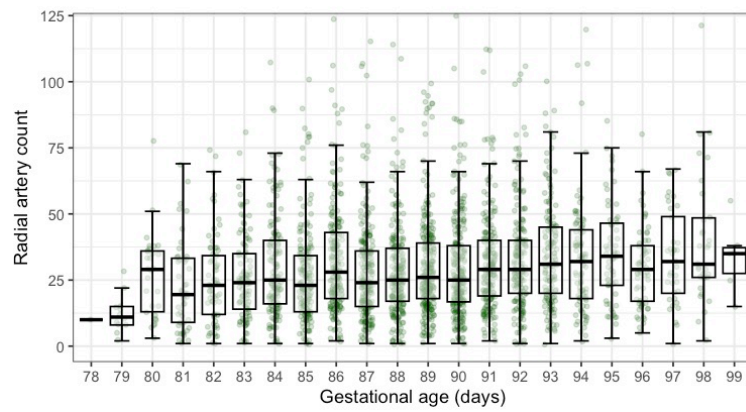
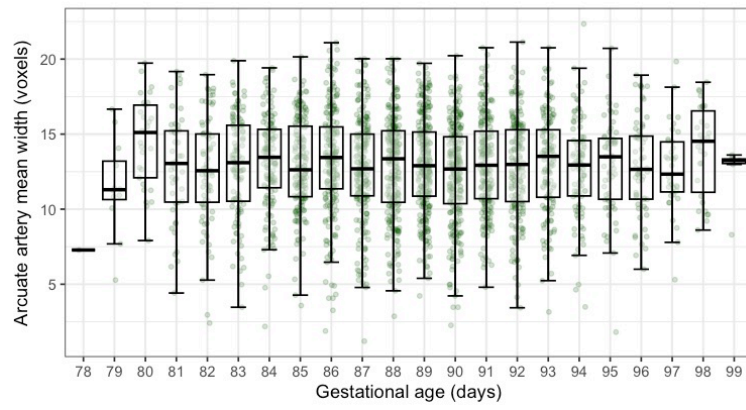
Model	10% FPR			
	Sens	Spec	PPV	NPV
FGR1	0.35 (0.26–0.45)	0.90 (0.88–0.91)	0.13 (0.09–0.17)	0.97 (0.96–0.98)
FGR2	0.23 (0.16–0.33)	0.90 (0.89–0.91)	0.09 (0.06–0.13)	0.96 (0.96–0.97)
FGR3	0.40 (0.30–0.50)	0.89 (0.88–0.90)	0.13 (0.10–0.18)	0.97 (0.96–0.98)
FGR4	0.37 (0.28–0.47)	0.89 (0.88–0.90)	0.13 (0.09–0.17)	0.97 (0.96–0.98)

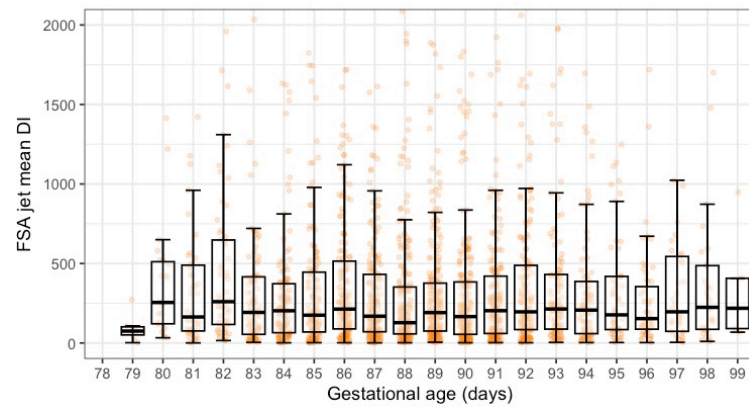
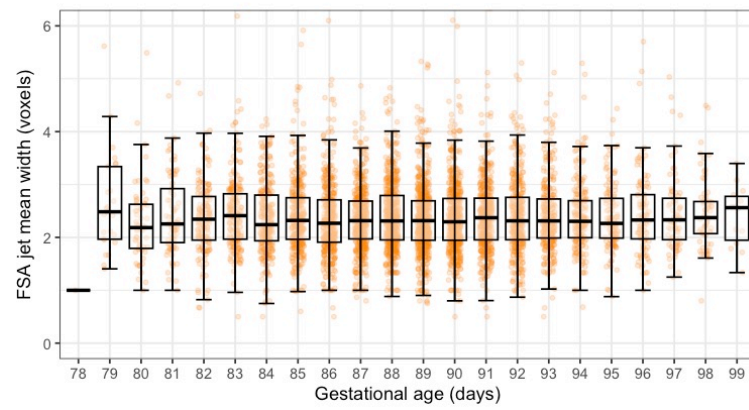
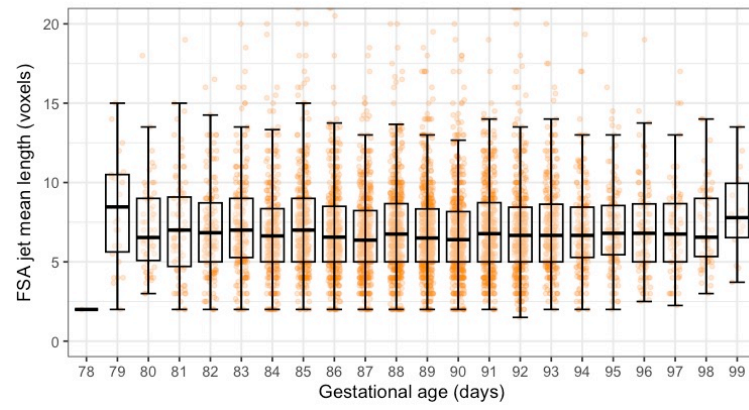
Model	30% FPR			
	Sens	Spec	PPV	NPV
FGR1	0.61 (0.51–0.71)	0.70 (0.68–0.72)	0.08 (0.06–0.10)	0.98 (0.97–0.98)
FGR2	0.48 (0.38–0.58)	0.70 (0.68–0.72)	0.06 (0.05–0.08)	0.97 (0.96–0.98)
FGR3	0.59 (0.49–0.69)	0.70 (0.68–0.72)	0.08 (0.06–0.10)	0.98 (0.97–0.98)
FGR4	0.67 (0.57–0.76)	0.70 (0.68–0.72)	0.09 (0.07–0.11)	0.98 (0.97–0.99)

Model	50% FPR			
	Sens	Spec	PPV	NPV
FGR1	0.84 (0.76–0.91)	0.49 (0.47–0.51)	0.07 (0.05–0.08)	0.99 (0.98–0.99)
FGR2	0.62 (0.52–0.72)	0.50 (0.48–0.53)	0.05 (0.04–0.06)	0.97 (0.96–0.98)
FGR3	0.72 (0.62–0.80)	0.50 (0.48–0.52)	0.06 (0.05–0.07)	0.98 (0.97–0.98)
FGR4	0.73 (0.63–0.81)	0.51 (0.49–0.53)	0.06 (0.05–0.07)	0.98 (0.97–0.99)

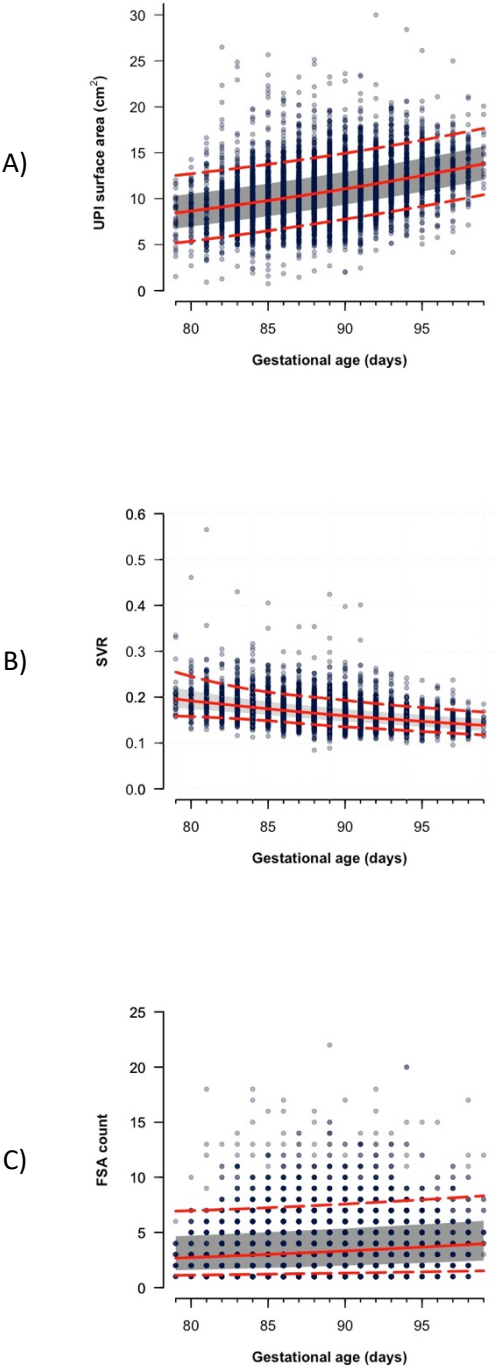
Appendix 8: Faceted boxplots exploring the changing distribution of data for novel placental markers with advancing gestational age







Appendix 9: Centile charts derived from LMS modelling of A) uteroplacental interface surface area, B) surface volume ratio, and C) functional spiral artery count



Appendix 10: Comparison of demographic characteristics and outcomes between processing success and processing failure subgroups with respect to novel placental markers

A) Comparison of processing success and failure subgroups for UPI surface area, UPI index, and SVR

Variable	Processing successful	Processing failed	p
	Median (IQR) or N (%) N = 6149	Median (IQR) or N (%) N = 97	
Age	33.11 (29.88–35.99)	33.89 (30.20–36.53)	0.544
Height	166 (161–170)	166 (162–171)	0.485
Weight	68 (60–78)	68 (62–83)	0.411
BMI	24.93 (22.00–28.60)	25.13 (22.06–30.85)	0.510
Nulliparity	3088 (50.22)	43 (44.33)	0.294
Smoking	185 (3.01)	1 (1.03)	0.372
Ethnicity			
Black	522 (8.49)	13 (13.40)	0.579
East Asian	112 (1.82)	2 (2.06)	
South Asian	470 (7.64)	7 (7.22)	
White	4782 (77.77)	71 (73.20)	
Mixed/other	263 (4.28)	4 (4.12)	
Gestational age at birth (days)	277 (272–284)	276 (272–283)	0.655
PE	224 (3.64)	6 (6.19)	0.295
PIH	268 (4.36)	1 (1.03)	0.130
SGA	625 (10.19)	17 (17.53)	0.029
FGR	346 (5.65)	16 (16.49)	<0.001

B) Comparison of processing success and failure subgroups for arcuate artery count

Variable	Processing successful	Processing failed	p
	Median (IQR) or N (%) N = 2560	Median (IQR) or N (%) N = 3686	
Age	32.98 (29.70–35.78)	33.20 (29.98–36.13)	0.080
Height	165 (161–170)	166 (161–170)	0.187
Weight	67 (60–77)	69 (61–80)	<0.001
BMI	24.39 (22.00–28.00)	25.00 (22.17–29.00)	<0.001
Nulliparity	1169 (45.66)	1962 (53.23)	<0.001
Smoking	69 (2.70)	117 (3.17)	0.308
Ethnicity			
Black	178 (6.95)	357 (9.69)	<0.001
East Asian	50 (1.95)	64 (1.74)	
South Asian	226 (8.83)	251 (6.81)	
White	2008 (78.44)	2845 (77.18)	
Mixed/other	98 (3.83)	169 (4.58)	
Gestational age at birth (days)	277 (273–283)	277 (272–284)	0.024
PE	88 (3.44)	142 (3.85)	0.432
PIH	119 (4.65)	150 (4.07)	0.294
SGA	245 (9.63)	397 (10.77)	0.156
FGR	136 (5.35)	226 (6.14)	0.208

C) Comparison of processing success and failure subgroups for radial artery count and radial artery width

Variable	Processing successful	Processing failed	p
	Median (IQR) or N (%) N = 2571	Median (IQR) or N (%) N = 3675	
Age	32.98 (29.69–35.78)	33.20 (29.99–36.12)	0.066
Height	165 (161–170)	166 (161–170)	0.218
Weight	67 (60–77)	69 (61–80)	<0.001
BMI	24.39 (22.00–28.00)	25.00 (22.15–29.00)	<0.001
Nulliparity	1175 (45.70)	1956 (53.22)	<0.001
Smoking	70 (2.72)	116 (3.16)	0.359
Ethnicity			
Black	180 (7.00)	355 (9.66)	<0.001
East Asian	51 (1.98)	63 (1.71)	
South Asian	225 (8.75)	252 (6.86)	
White	2016 (78.41)	2837 (77.20)	
Mixed/other	99 (3.85)	168 (4.57)	
Gestational age at birth (days)	277 (273–283)	277 (272–284)	0.017
PE	90 (3.50)	140 (3.81)	0.571
PIH	119 (4.63)	150 (4.08)	0.323
SGA	248 (9.70)	394 (10.72)	0.208
FGR	138 (5.40)	224 (6.10)	0.267

D) Comparison of processing success and failure subgroups for functional spiral artery jet count

Variable	Processing successful	Processing failed	p
	Median (IQR) or N (%) N = 5555	Median (IQR) or N (%) N = 691	
Age	33.13 (29.84–35.97)	33.10 (30.13–36.20)	0.688
Height	165 (161–170)	166 (162–170)	0.464
Weight	68 (60–78)	70 (62–83)	<0.001
BMI	24.78 (22.00–28.28)	25.59 (22.60–30.34)	<0.001
Nulliparity	2791 (50.24)	340 (49.20)	0.635
Smoking	164 (2.95)	22 (3.18)	0.827
Ethnicity			
Black	465 (8.37)	70 (10.13)	0.619
East Asian	101 (1.82)	13 (1.88)	
South Asian	424 (7.63)	53 (7.67)	
White	4325 (77.86)	528 (76.41)	
Mixed/other	240 (4.32)	27 (3.91)	
Gestational age at birth (days)	277 (272–284)	276 (272–283)	0.055
PE	199 (3.58)	31 (4.49)	0.279
PIH	242 (4.36)	27 (3.91)	0.653
SGA	560 (10.11)	82 (11.87)	0.171
FGR	311 (5.62)	51 (7.38)	0.075

E) Comparison of processing success and failure subgroups for functional spiral artery dispersion index

Variable	Processing successful	Processing failed	p
	Median (IQR) or N (%) N = 2300	Median (IQR) or N (%) N = 3946	
Age	33.03 (29.64–35.76)	33.17 (30.00–36.13)	0.066
Height	165 (161–170)	166 (161–170)	0.704
Weight	67 (59–77)	69 (61–80)	<0.001
BMI	24.20 (21.98–28.00)	25.00 (22.23–29.00)	<0.001
Nulliparity	1053 (45.78)	2078 (52.66)	<0.001
Smoking	64 (2.78)	122 (3.09)	0.538
Ethnicity			
Black	154 (6.70)	381 (9.66)	
East Asian	47 (2.04)	67 (1.70)	
South Asian	197 (8.57)	280 (7.10)	<0.001
White	1813 (78.83)	3040 (77.04)	
Mixed/other	89 (3.87)	178 (4.51)	
Gestational age at birth (days)	277 (273–283)	277 (272–284)	0.051
PE	82 (3.57)	148 (3.75)	0.762
PIH	107 (4.65)	162 (4.11)	0.334
SGA	224 (9.80)	418 (10.59)	0.345
FGR	127 (5.56)	235 (5.96)	0.550