Ultrasound Obstet Gynecol 2020; **56**: 541–548
Published online in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.22175

Competing-risks model for prediction of small-for-gestational-age neonate from maternal characteristics and serum pregnancy-associated plasma protein-A at 11–13 weeks' gestation

I. PAPASTEFANOU[1], D. WRIGHT[2], A. SYNGELAKI[1], M. LOLOS[1], K. ANAMPOUSI[1] and K. H. NICOLAIDES[1]

[1]*Fetal Medicine Research Institute, King's College Hospital, London, UK;* [2]*Institute of Health Research, University of Exeter, Exeter, UK*

KEYWORDS: Bayes' theorem; fetal growth restriction; first-trimester screening; PAPP-A; SGA; survival model

CONTRIBUTION

What are the novel findings of this work?
This study describes a new competing-risks model based on a combination of maternal characteristics and medical history with serum pregnancy-associated plasma protein-A (PAPP-A) at 11–13 weeks' gestation for prediction of a small-for-gestational-age (SGA) neonate. PAPP-A likelihood was expressed as a continuous function of both gestational age at delivery and birth-weight *Z*-score in the same model.

What are the clinical implications of this work?
Addition of serum PAPP-A improves the performance of screening for a SGA neonate achieved by maternal factors alone and demonstrates the methodology for incorporation of further biomarkers into a single model that can be used numerous times during the course of pregnancy to predict SGA of any severity of smallness and degree of prematurity.

ABSTRACT

Objectives *To develop a continuous likelihood model for pregnancy-associated plasma protein-A (PAPP-A), in the context of a new competing-risks model for prediction of a small-for-gestational-age (SGA) neonate, and to compare the predictive performance of the new model for SGA to that of previous methods.*

Methods *This was a prospective observational study of 60 875 women with singleton pregnancy undergoing routine ultrasound examination at 11 + 0 to 13 + 6 weeks' gestation. The dataset was divided randomly into a training dataset and a test dataset. The training dataset was used for PAPP-A likelihood model development. We used Bayes' theorem to combine the previously developed prior model for the joint Gaussian distribution of gestational age (GA) at delivery and birth-weight Z-score with the PAPP-A likelihood to obtain a posterior distribution. This patient-specific posterior joint Gaussian distribution of GA at delivery and birth-weight Z-score allows risk calculation for SGA defined in terms of different birth-weight percentiles and GA. The new model was validated internally in the test dataset and we compared its predictive performance to that of the risk-scoring system of the UK National Institute for Health and Care Excellence (NICE) and that of logistic regression models for different SGA definitions.*

Results *PAPP-A has a continuous association with both birth-weight Z-score and GA at delivery according to a folded-plane regression. The new model, with the addition of PAPP-A, was equal or superior to several logistic regression models. The new model performed well in terms of risk calibration and consistency across different GAs and birth-weight percentiles. In the test dataset, at a false-positive rate of about 30% using the criteria defined by NICE, the new model predicted 62.7%, 66.5%, 68.1% and 75.3% of cases of a SGA neonate with birth weight $< 10^{th}$ percentile delivered at < 42, < 37, < 34 and < 30 weeks' gestation, respectively, which were significantly higher than the respective values of 46.7%, 55.0%, 55.9% and 52.8% achieved by application of the NICE guidelines.*

Correspondence to: Prof. K. H. Nicolaides, Fetal Medicine Research Institute, King's College Hospital, 16–20 Windsor Walk, Denmark Hill, London SE5 8BB, UK (e-mail: kypros@fetalmedicine.com)

Accepted: 24 July 2020

Conclusions *Using Bayes' theorem to combine PAPP-A measurement data with maternal characteristics improves the prediction of SGA and performs better than logistic regression or NICE guidelines, in the context of a new competing-risks model for the joint distribution of birth-weight Z-score and GA at delivery. © 2020 International Society of Ultrasound in Obstetrics and Gynecology*

INTRODUCTION

Small-for-gestational-age (SGA) neonates are at increased risk of adverse perinatal outcome and development of metabolic and cardiovascular diseases in adult life[1–5]. The monitoring and the decision for delivery of pregnancies suspected for SGA is the subject of guidelines issued by relevant national societies[6]. The optimal way to identify SGA fetuses is debatable[7]. The mainstream approach is to recognize a high-risk group for SGA by application of a scoring system. For example, in the UK, according to guidelines of the National Institute for Health and Care Excellence (NICE), women should be considered to be at high risk if they have any one major risk factor or any three minor risk factors[8]. The major risk factors include maternal age >40 years, smoking, previous SGA baby or stillbirth, chronic hypertension, diabetes with vascular disease, renal impairment, antiphospholipid syndrome and serum pregnancy-associated plasma protein-A (PAPP-A) <0.4 multiples of the median (MoM). Minor risk factors include nulliparity, maternal age ≥ 35 years, body mass index <20 or 25–34.9 kg/m^2, conception by *in-vitro* fertilization, previous pre-eclampsia (PE) and interpregnancy interval <6 or ≥ 60 months. Although this approach is relatively simple to perform, it does not provide patient-specific risks and has uncertain performance in predicting a SGA neonate. Another approach is to use probabilistic models treating SGA as a binary outcome and applying logistic regression to develop different models for different definitions of SGA, such as birth weight $<10^{th}$, $<5^{th}$ or $<3^{rd}$ percentile born at <40 or <37 or <34 weeks' gestation. These models use maternal characteristics and medical history alone or in combination with biomarkers to quantify the individual patient-specific risk for SGA, rather than just classifying women into high- and low-risk groups[9–12]. However, each time we want to predict a different SGA definition and/or add a new biomarker, the whole model must be refitted.

We have demonstrated an alternative method for prediction of SGA, similar to the competing-risks model in the assessment of the risk for PE[13–16]. This new method is based on a continuous personalized joint bivariate Gaussian distribution of gestational age (GA) at delivery and *Z*-score of birth weight, that allows risk calculation for any desired SGA definition[17]. An important merit of the new model is the ability to easily include biomarkers, according to Bayes' rule.

The objective of this study was to expand a new continuous history model for the prediction of a SGA neonate, with the Bayesian incorporation of PAPP-A. We assessed the predictive performance of the new model and we compared it with logistic regression models and the application of NICE guidelines.

METHODS

Study population and design

This was a prospective screening study for adverse obstetric outcomes in women attending for their routine first hospital visit in pregnancy at King's College Hospital, London and Medway Maritime Hospital, Gillingham, UK, between March 2006 and December 2016. At this visit, at $11+0$ to $13+6$ weeks' gestation, we recorded maternal characteristics and medical history, performed combined screening for aneuploidy[18] and measured serum concentration of PAPP-A (DELFIA® Xpress system, PerkinElmer Life and Analytical Sciences, Waltham, MA, USA). GA was determined by the measurement of fetal crown–rump length[19]. The participants gave written informed consent for the study, which was approved by the UK National Health Service Research Ethics Committee. We included singleton pregnancies that resulted in a non-malformed liveborn or stillborn neonate at ≥ 24 weeks' gestation. Pregnancies with aneuploidy and major fetal abnormality and those ending in termination, miscarriage or fetal death at <24 weeks' gestation, were excluded from the dataset.

Outcome measures

Data on pregnancy outcome were collected from hospital maternity records or the general medical practitioners of the women. The outcome measures of the study were birth of a neonate at or below different thresholds of birth-weight percentile for different cut-offs of GA at delivery. The Fetal Medicine Foundation fetal and neonatal population weight charts were used to convert birth weight to percentiles and *Z*-scores[20].

Statistical analysis

We converted serum concentrations of PAPP-A to MoM values, as described previously[18].

Model development

The new model is based on a personalized joint distribution of GA at delivery and birth-weight *Z*-score. The two elements that defined this distribution were, first, the prior distribution determined by maternal factors and, second, the likelihood of the biomarker values. The prior distribution was obtained using the history model, as described previously[17]. We expressed $\log_{10}$ PAPP-A MoM likelihood conditionally to birth weight and GA at delivery. We adopted a two-dimensional extension of the

broken-stick regression, a folded-plane regression. The mean $\log_{10}$ PAPP-A MoM was a linear combination of birth-weight *Z*-score and GA at delivery, until it reaches zero level, and beyond a break line the mean was presumed to be constant and equal to zero. We used Bayes' rule to combine the prior and the likelihood to obtain a posterior joint distribution for each pregnancy, that can be used to compute risks for different cut-offs of birth-weight *Z*-score and GA at delivery. Residual diagnostics were used to assess the fit of the model.

Training and test datasets

Data were partitioned randomly into a training dataset of 30 438 cases and a test dataset of 30 437 cases. The training data were used for model fitting and the model was then assessed on the test dataset for the purpose of internal validation.

Predictive performance

We examined the predictive performance of the new model by means of detection rate (DR) of a SGA neonate of different severities ($< 10^{th}$ and $< 3^{rd}$ percentiles) at different gestational age cut-offs (< 42, < 37, < 34 and < 30 weeks), at fixed false-positive rates (FPR) of 5%, 10% and 20%. Calibration intercept and slope using logistic regression analysis of outcome incidence against the logit of the respective risks were computed.

Comparison with previous definitions of SGA and logistic regression models

A series of logistic regression models to predict SGA ($< 10^{th}$ and $< 3^{rd}$ percentiles for GA at birth < 42, < 37, < 34 and < 30 weeks) were fitted and validated in the test dataset.

Comparison of performance of new model with that of NICE guidelines

We obtained the risk cut-off that gave the same FPR as that of the NICE guidelines[8], for all women and separately for nulliparous and parous women, and used McNemar's test to assess the differences in DRs between the new model and the risk-scoring system proposed by the NICE guidelines[8].

Model fitting was carried out within a Bayesian framework using Markov chain Monte Carlo[21]. The statistical software package R was also used for data analyses[22].

RESULTS

Datasets

The study population included 60 875 singleton pregnancies. The maternal and pregnancy characteristics in the training and validation datasets are given in Table 1. The two datasets had a similar distribution for all variables and no significant differences were observed.

Likelihood function for PAPP-A

A folded-plane regression model for the mean $\log_{10}$ PAPP-A MoM was fitted to the training dataset. The inferences for the parameters are presented in Table 2. Residual diagnostics revealed satisfactory fitting for the likelihood model in the test dataset. The folded-plane regression is depicted in a three-dimensional representation in Figure 1. It is obvious that the PAPP-A levels increase with increasing birth-weight *Z*-score and GA at delivery until the break line, and beyond this threshold, defined by the model, PAPP-A remains constant

Table 1 Maternal and pregnancy characteristics in training and test datasets

Variable	*Training dataset (n = 30 438)*	*Test dataset (n = 30 437)*
Maternal age (years)	31.0 (26.5–34.8)	31.1 (26.6–34.8)
Maternal weight (kg)	67.1 (59.4–78.3)	67.1 (59.4–78.1)
Maternal height (cm)	165 (160–169)	165 (160–169)
BMI (kg/m^2)	24.7 (22.0–28.8)	24.8 (22.1–28.7)
GA (weeks)	12.7 (12.3–13.1)	12.7 (12.3–13.1)
Racial origin		
White	22 498 (73.9)	22 458 (73.8)
Black	5138 (16.9)	5251 (17.3)
South Asian	1392 (4.6)	1332 (4.4)
East Asian	626 (2.1)	628 (2.1)
Mixed	784 (2.6)	768 (2.5)
Conception		
Natural	29 433 (96.7)	29 469 (96.8)
Ovulation induction	244 (0.8)	249 (0.8)
In-vitro fertilization	761 (2.5)	719 (2.4)
Medical history		
Chronic hypertension	435 (1.4)	410 (1.3)
Diabetes mellitus	266 (0.9)	294 (1.0)
SLE/APS	60 (0.2)	62 (0.2)
Cigarette smoker	2828 (9.3)	2940 (9.7)
Family history of PE	1206 (4.0)	1187 (3.9)
Parity		
Nulliparous	14 134 (46.4)	14 177 (46.6)
Parous with previous PE or SGA $< 10^{th}$ percentile	2743 (9.0)	2783 (9.1)
Parous with previous SGA $< 10^{th}$ percentile	2304 (7.6)	2362 (7.8)
Parous with previous PE and SGA $< 10^{th}$ percentile	247 (0.8)	232 (0.8)
Interpregnancy interval (years)	3.0 (2.0–4.9)	3.0 (2.0–4.9)
GA at delivery of last pregnancy (weeks)	40.0 (39.0–40.0)	40.0 (39.0–40.0)

Data are given as median (interquartile range) or *n* (%). APS, antiphospholipid syndrome; BMI, body mass index; GA, gestational age; PE, pre-eclampsia; SGA, small-for-gestational age; SLE, systemic lupus erythematosus.

Table 2 Fitted folded-plane regression model for mean $\log_{10}$ pregnancy-associated plasma protein-A multiples of the median conditional to birth-weight *Z*-score and gestational age at delivery (GA)

Term	*Estimate (95% credibility interval)*	*SD*
Intercept	0.021158030 (0.01083000–0.03231025)	0.00535788
Birth-weight *Z*-score	0.042875322 (0.03698000–0.04895000)	0.00300972
GA – 40	0.016274441 (0.01120000–0.02234025)	0.00283057
$(GA - 40)^2$	0.001214426 (0.00059803–0.00207300)	0.00037523
Residual SD	0.240294492 (0.23840000–0.24220000)	0.00097587

Posterior mean, 95% credibility interval and SD for each parameter are presented.

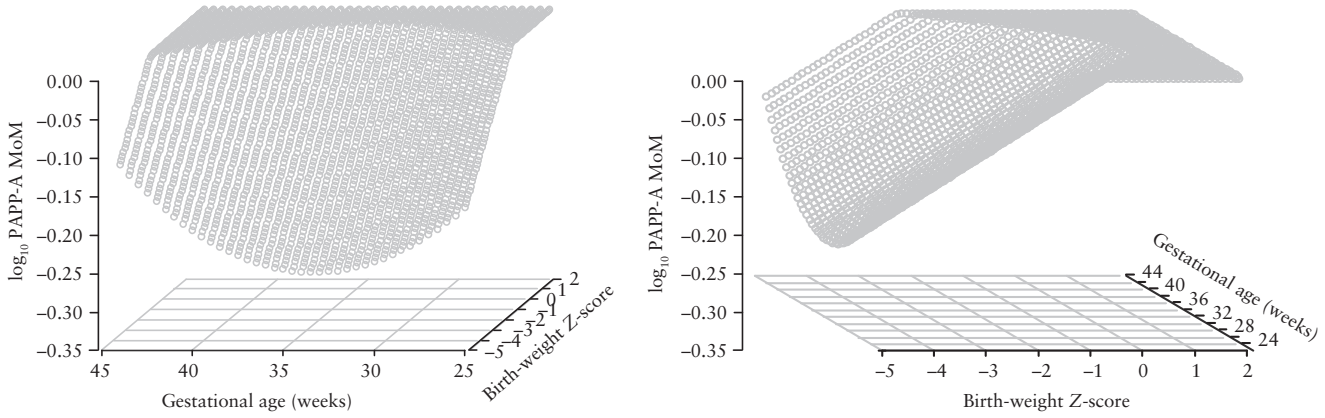

Figure 1 Three-dimensional representation of folded-plane regression for pregnancy-associated plasma protein-A (PAPP-A) multiples of the median (MoM) likelihood model from two different angles.

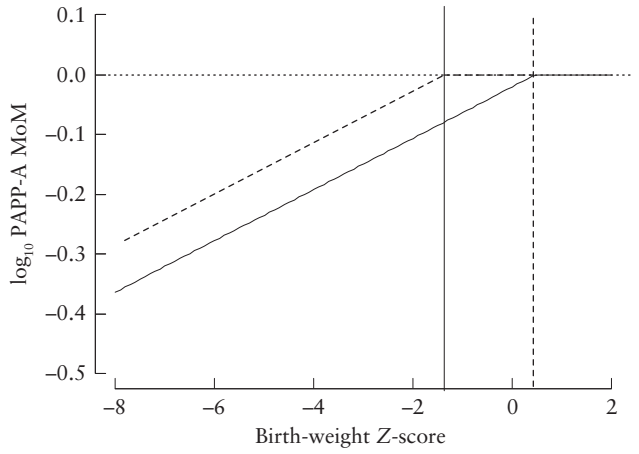

Figure 2 Folded-plane likelihood model reduced to two-dimensional graph on which broken-stick model for pregnancy-associated plasma protein-A (PAPP-A) multiples of the median (MoM) conditional to birth-weight *Z*-score is depicted. Break point is function of gestational age. Dashed line corresponds to 40 weeks and solid line to 30 weeks.

and zero. Figure 2 illustrates in a two-dimensional demonstration how the likelihood works. Essentially, a broken stick is fitted conditionally to birth-weight *Z*-score and the model allows the break point to be a function of GA at delivery. Figure 3 shows the joint distribution of birth-weight *Z*-score and GA at delivery updated by the addition of PAPP-A in a high-risk and low-risk case. For the high-risk case, the contour lines gravitate towards earlier gestations and lower birth weights.

Model evaluation

The prediction of several SGA definitions at fixed FPRs is presented in Table 3. The prediction improved gradually for earlier gestations and increasing severity of SGA. The prediction was also better for parous women (Table 4). As expected, the DRs were mostly lower in the test dataset (Table 3). Overall, we found that the agreement between the predicted risks by the new model for SGA and the observed incidence for different SGA definitions was good (Table 5). The calibration indices were similar in the training and test datasets; therefore, we would expect realistic risks on clinical application of the model.

Comparison of performance of new model with that of NICE guidelines

The predictive performance of the competing-risks model was superior to that of the scoring system proposed by the NICE guidelines (Table 6). At a FPR of about 30%, as defined by NICE, the new model predicted 62.7%, 66.5%, 68.1% and 75.3% of cases of a SGA neonate with birth weight $< 10^{th}$ percentile delivered at < 42, < 37, < 34 and < 30 weeks' gestation, respectively, compared to the

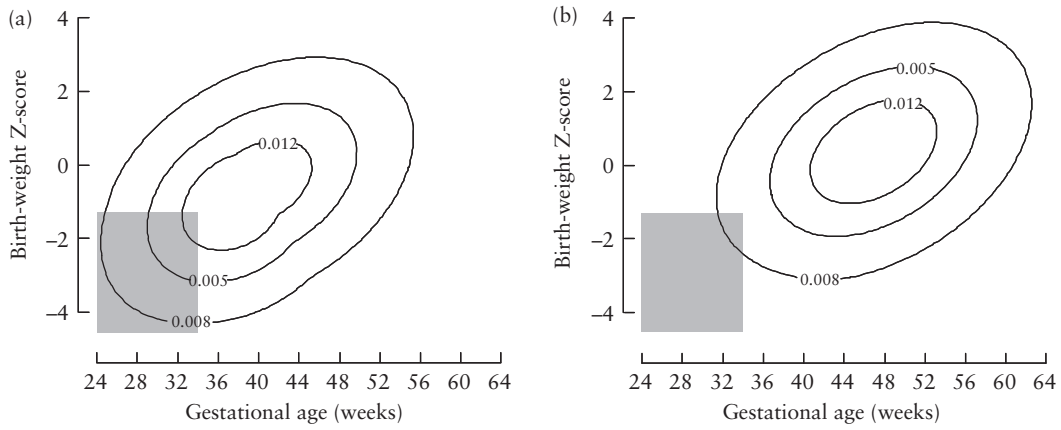

Figure 3 Contour plots of joint Gaussian distribution of birth-weight *Z*-scores and gestational age at delivery according to maternal factors and pregnancy-associated plasma protein-A in high-risk (a) and low-risk (b) case. Shaded area corresponds to risk of delivery before 34 weeks' gestation with birth weight below 10th percentile.

Table 3 Comparison of performance of screening by maternal factors and pregnancy-associated plasma protein-A (PAPP-A) in prediction of small-for-gestational-age neonate with birth weight (BW) < 10th or < 3rd percentile, for different gestational-age cut-offs at delivery, between new model and logistic regression models, in training and test datasets

			DR (%) at FPR of:					
	AUC		*5%*		*10%*		*20%*	
Outcome measure	*Training*	*Test*	*Training*	*Test*	*Training*	*Test*	*Training*	*Test*
Delivery < 42 weeks								
BW < 10th percentile								
New model (maternal factors)	0.7212	0.7200	18.3	18.2	29.9	30.2	48.2	48.4
New model (maternal factors + PAPP-A)	0.7365	0.7396	20.6	20.4	32.6	34.3	50.7	51.3
Logistic regression	0.7388	0.7405	20.4	21.4	32.6	34.3	51.3	52.1
BW < 3rd percentile								
New model (maternal factors)	0.7465	0.7388	21.9	20.6	33.1	34.3	51.3	51.2
New model (maternal factors + PAPP-A)	0.7716	0.7643	25.1	25.1	37.5	38.5	56.9	56.3
Logistic regression	0.7680	0.7629	24.9	25.0	37.4	38.1	56.7	55.5
Delivery < 37 weeks								
BW < 10th percentile								
New model (maternal factors)	0.7368	0.7039	23.6	20.0	34.7	30.0	50.5	46.2
New model (maternal factors + PAPP-A)	0.7680	0.7512	26.9	23.4	39.9	36.5	57.1	55.7
Logistic regression	0.7687	0.7464	26.9	23.1	40.3	33.5	58.6	53.7
BW < 3rd percentile								
New model (maternal factors)	0.7536	0.7211	25.2	20.8	36.0	31.7	53.6	48.4
New model (maternal factors + PAPP-A)	0.7919	0.7736	31.7	26.1	44.1	41.3	62.6	59.1
Logistic regression	0.7994	0.7716	33.6	25.7	47.6	37.0	63.8	57.2
Delivery < 34 weeks								
BW < 10th percentile								
New model (maternal factors)	0.7410	0.7111	25.6	21.7	34.0	32.3	52.1	46.1
New model (maternal factors + PAPP-A)	0.7700	0.7603	29.4	27.2	41.6	39.0	58.8	58.3
Logistic regression	0.7693	0.7438	31.9	24.4	44.9	33.1	59.2	52.3
BW < 3rd percentile								
New model (maternal factors)	0.7558	0.7190	24.9	20.9	32.2	31.3	49.7	49.3
New model (maternal factors + PAPP-A)	0.7710	0.7739	29.4	28.9	40.7	39.3	57.6	60.2
Logistic regression	0.7910	0.7563	33.9	24.4	44.6	34.8	65.5	53.2
Delivery < 30 weeks								
BW < 10th percentile								
New model (maternal factors)	0.7407	0.7317	26.4	22.5	31.9	38.2	46.2	48.3
New model (maternal factors + PAPP-A)	0.7653	0.7807	33.0	24.7	39.6	41.6	55.0	60.7
Logistic regression	0.7515	0.7725	34.1	21.3	43.9	31.5	59.3	61.8
BW < 3rd percentile								
New model (maternal factors)	0.7431	0.7115	27.9	22.5	33.8	35.2	45.6	45.1
New model (maternal factors + PAPP-A)	0.7708	0.7684	30.9	25.4	39.7	38.0	54.4	57.8
Logistic regression	0.7565	0.7574	30.9	19.7	42.6	38.0	58.8	57.8

AUC, area under the receiver-operating-characteristics curve; DR, detection rate; FPR, false-positive rate.

Table 4 Comparison of performance of screening for small-for-gestational-age (SGA) neonate with birth weight < 10th percentile, for different gestational-age (GA) cut-offs at delivery, between new model and risk-scoring system based on National Institute for Health and Care Excellence (NICE) guidelines, in test dataset, overall and according to parity

					Detection rate (% (95% CI))		
GA at delivery	*Total* (n)	*SGA* (n (%))	*FPR (%)*	*Risk cut-off (probability (1/*N*))*	*NICE guidelines*	*New model*	P
< 42 weeks							
Total	30 437	3959 (13.00)	29.40	0.05535072 (1/18)	46.7 (45.1–48.2)	62.7 (62.2–65.2)	< 0.0001
Nulliparous	14 177	2239 (15.79)	22.55	0.06402645 (1/16)	34.0 (32.1–36.0)	55.4 (53.3–57.4)	< 0.0001
Parous	16 260	1720 (10.58)	35.03	0.03545928 (1/28)	63.1 (60.8–65.4)	74.8 (72.7–76.8)	< 0.0001
< 37 weeks							
Total	30 437	671 (2.21)	31.11	0.01067656 (1/94)	55.0 (51.2–58.8)	66.5 (62.9–70.0)	< 0.0001
Nulliparous	14 177	353 (2.49)	23.90	0.01543537 (1/65)	42.5 (37.3–47.7)	51.3 (46.1–56.5)	0.00406
Parous	16 260	318 (1.96)	37.38	0.006030673 (1/166)	68.9 (63.8 – 74.0)	79.6 (75.1–84.0)	0.00044
< 34 weeks							
Total	30 437	254 (0.84)	31.44	0.007416389 (1/135)	55.9 (49.8–62.0)	68.1 (62.4–73.8)	0.00053
Nulliparous	14 177	138 (0.97)	27.51	0.009509288 (1/105)	44.2 (35.9–52.5)	53.6 (45.3–61.9)	0.03737
Parous	16 260	116 (0.71)	37.77	0.004490741 (1/223)	69.8 (61.5–78.2)	82.8 (75.9–89.6)	0.01481
< 30 weeks							
Total	30 437	89 (0.29)	31.58	0.001497970 (1/668)	52.8 (42.4–63.2)	75.3 (66.3–84.2)	0.00033
Nulliparous	14 177	42 (0.30)	27.39	0.001912714 (1/523)	28.6 (14.9–42.2)	45.2 (30.2–60.3)	0.03481
Parous	16 260	47 (0.29)	37.89	0.0008853714 (1/1129)	74.5 (62.0–86.9)	89.4 (80.6–98.2)	0.03481

False-positive rate (FPR) in each outcome group is that derived from NICE guidelines. Risk cut-off for new model in each outcome group is that corresponding to FPR derived by NICE guidelines. McNemar's test was used to compare detection rates between new model and NICE guidelines.

Table 5 Calibration study for new model in prediction of small-for-gestational-age neonate with birth weight (BW) < 10th or < 3rd percentile, for different gestational-age (GA) cut-offs at delivery, by maternal factors and pregnancy-associated plasma protein-A, in training and test datasets

	BW < 10th percentile		*BW < 3rd percentile*	
GA at delivery	*Slope*	*Intercept*	*Slope*	*Intercept*
< 42 weeks				
Training dataset	0.97290	1.06168	0.95722	0.68681
Test dataset	1.00301	1.0469	0.95649	0.7171
< 37 weeks				
Training dataset	0.93390	0.10714	0.91686	0.14773
Test dataset	0.88789	0.12846	0.8718	0.26218
< 34 weeks				
Training dataset	0.89808	−0.06632	0.82473	0.07248
Test dataset	0.85072	0.004051	0.82710	0.21143
< 30 weeks				
Training dataset	0.80476	0.4420	0.78275	0.5094
Test dataset	0.78279	0.4234	0.73822	0.56350

respective values of 46.7%, 55.0%, 55.9% and 52.8% achieved by application of the NICE guidelines (Table 4).

Comparison of performance of new model with that of logistic regression models

The predictive performance of the new model for SGA < 10th and < 3rd percentiles for gestational ages at birth < 42, < 37, < 34 and < 30 weeks, for fixed FPRs, was equal or superior to that of several logistic regression models (Table 3). The process of internal validation demonstrated that the new model is more stable with superior performance for preterm SGA (Table 3). The logistic regression models showed a large drop in their discriminative ability in the test dataset, especially for preterm SGA.

DISCUSSION

Principal findings

This study demonstrates that SGA is one condition, described by the continuous combination of GA at delivery and birth-weight *Z*-score (Figure 3). A single model can be used for the prediction of any SGA definition. This oneness is also reflected in the distribution of PAPP-A, which is a continuous function of both GA at delivery and birth-weight *Z*-score (Figures 1 and 2).

Table 6 Variables used in National Institute for Health and Care Excellence (NICE) scoring system and new competing-risks model for prediction of small-for-gestational-age (SGA) neonate

NICE guidelines	*New competing-risks model*
Race not included	Race included
Minor risk factors (three or more)	
Maternal age ≥ 35 years	Maternal age examined and not included
Conception by IVF	Conception by IVF included
Nulliparous	Parity included as protective factor
BMI < 20 kg/m^2	Maternal weight and height included as continuous variables
BMI 25–34.9 kg/m^2	Maternal weight and height included as continuous variables
Smoker 1–10 cigarettes per day	Smoking status included
Low fruit intake prepregnancy	Fruit intake not available
Previous pre-eclampsia	Previous pre-eclampsia included
Pregnancy interval < 6 months	Pregnancy interval included as continuous variable
Pregnancy interval ≥ 60 months	Pregnancy interval included as continuous variable
Major risk factors (one or more)	
Maternal age > 40 years	Maternal age examined and not included
Smoker ≥ 11 cigarettes per day	Smoking status included
Paternal SGA	Paternal SGA not available
Maternal SGA	Maternal SGA not available
Previous SGA baby	*Z*-scores of birth weight in previous pregnancy included as continuous variable
Cocaine use	Cocaine use not available
Daily vigorous exercise	Exercise not available
Previous stillbirth	Previous stillbirth included
Chronic hypertension	Chronic hypertension included
Diabetes with vascular disease	Any type of diabetes included
Renal impairment	Renal impairment not included
APS	APS included as SLE and/or APS
Heavy bleeding similar to menses	Bleeding not included
PAPP-A < 0.4 MoM	PAPP-A included as continuous likelihood

APS, antiphospholipid syndrome; BMI, body mass index; IVF, *in-vitro* fertilization; MoM, multiples of the median; PAPP-A, pregnancy-associated plasma protein-A; SLE, systemic lupus erythematosus.

Therefore, a new rationale is introduced that overcomes the historical usage of fragmented outcomes and biomarker thresholds, leading to the erroneous conclusion that SGA consists of several different outcomes.

We confirm that PAPP-A is lower in SGA pregnancies. Our approach goes beyond the known observation that the larger the deviation of the biomarker the higher the risk for preterm SGA. The deviation from a normal value of 1 MoM depends continuously and simultaneously on the GA at delivery and birth-weight *Z*-score. The use of fixed thresholds does not capture efficiently the association between the biomarker and the risk for SGA. According to NICE guidelines, PAPP-A levels below 0.4 MoM are considered as a risk factor for SGA; consequently, pregnancies with low PAPP-A, but above 0.4 MoM, that still have a substantial risk for SGA, are considered screen negative. On the other hand, the risk that is attached to each pregnancy, when logistic regression models are used, is a continuous function of the biomarker levels for the whole range of the biomarkers values. However, the folded-plane regression that we fitted for PAPP-A proves that the continuous relationship between PAPP-A and birth-weight *Z*-score and GA at delivery continues until the level of 1 MoM. Therefore, the model is focused on the biomarker levels that are clinically relevant, in a continuous way. Practically, we may now predict SGA using a single model. A whole personalized probability distribution is now assigned to each pregnancy that allows risk calculation for any desired cut-offs for severity of smallness and degree of prematurity. This study also demonstrates the process of adding biomarkers in the same model by using Bayes' rule.

The predictive performance of the model for a SGA neonate is superior to the risk-scoring system recommended by NICE guidelines[8]. The performance of screening in the test dataset is essentially a metric of the actual clinical use of a prediction model. This process of internal validation showed that, first, a single model has better performance compared to a series of different logistic regression models, that were fitted separately for the different SGA definitions, and, second, the new model is more stable in contrast to the logistic regression approach, which has diminished discrimination in the validation dataset, especially for preterm cases. The calibration of the model is good and remained almost unchanged in the validation dataset. Therefore, valid risks are produced, enhancing an early risk stratification for SGA.

Comparison with previous studies

Previous first-trimester studies that aimed to predict delivery of a SGA neonate reported similar sensitivities compared to that achieved by the new model[9–12]. However, the predictive performance of the new approach is actually higher than that of previous models because our definition of SGA was based on the new Fetal Medicine Foundation birth-weight charts; these charts modeled efficiently the overrepresentation of preterm SGA pregnancies, and this has led to an increasing percentage of SGA for lower GA cut-offs[20]. Thus, we are predicting an outcome that is less extreme, compared to the previous definitions, and consequently more difficult to predict.

Strengths and limitations

The strengths of this study are, first, the large dataset and its prospective nature in accordance with an implemented screening program, second, use of a continuous folded-surface model that best describes the

distribution of PAPP-A, third, use of a joint model that allows estimation of patient-specific risks for any desired SGA definition, and, fourth, use of Bayes' rule in an update process that can be repeated numerous times during the course of pregnancy. We internally validated the model in terms of discrimination and calibration, to gain insight into what to expect in a real clinical scenario. However, external validation is needed to show the applicability of our results in other populations.

Implications for clinical practice

In the new era of precision medicine, we aim to use a personalized joint distribution of birth-weight *Z*-score and GA at delivery that can be altered sequentially in a Bayesian framework. The resultant posterior distribution can be translated to a risk for an infinite number of combinations for birth-weight *Z*-score and GA at delivery cut-offs. The new approach for the prediction of SGA expands our thinking to a continuous association between the biomarker levels and both the degree of prematurity and severity of smallness. This method leads to a unified perspective in SGA prediction and management that can be simultaneously tailored to each pregnancy and applied at any GA. We use a single model to obtain a probability distribution for a continuous joint outcome rather than arbitrarily categorizing it beforehand. This methodology may improve the allocation of resources and planning for antenatal visits. Also, adverse outcomes related to SGA could be attached to a continuous model, enhancing our understanding of the disease.

Conclusions

The distribution of PAPP-A aligns with the two elements of SGA: severity of smallness and degree of prematurity. This study provides more evidence that SGA is one joint continuous outcome. The methodology described is the benchmark of adding a new biomarker. This can be extended to more biomarkers and repeated many times during the course of pregnancy.

ACKNOWLEDGMENT

This study was supported by grants from The Fetal Medicine Foundation (Charity No: 1037116).

REFERENCES

1. Pilliod RA, Cheng YW, Snowden JM, Doss AE, Caughey AB. The risk of intrauterine fetal death in the small-for-gestational-age fetus. *Am J Obstet Gynecol* 2012; **207**: 318–416.
2. Trudell AS, Cahill AG, Tuuli MG, Macones GA, Odibo AO. Risk of stillbirth after 37 weeks in pregnancies complicated by small-for-gestational-age fetuses. *Am J Obstet Gynecol* 2013; **208**: 376.e1–7.
3. Bukowski R, Burgett AD, Gei A, Saade GR, Hankins GD. Impairment of fetal growth potential and neonatal encephalopathy. *Am J Obstet Gynecol* 2003; **188**: 1011–1015.
4. McIntyre S, Blair E, Badawi N, Keogh J, Nelson KB. Antecedents of cerebral palsy and perinatal death in term and late preterm singletons. *Obstet Gynecol* 2013; **122**: 869–877.
5. Barker DJ. Adult consequences of fetal growth restriction. *Clin Obstet Gynecol* 2006; **49**: 270–283.
6. McCowan LM, Figueras F, Anderson NH. Evidence-based national guidelines for the management of suspected fetal growth restriction: comparison, consensus, and controversy. *Am J Obstet Gynecol* 2018; **218**: S855–S868.
7. Ciobanu A, Rouvali, A, Syngelaki A, Akolekar R, Nicolaides KH. Prediction of small for gestational age neonates: Screening by maternal factors, fetal biometry and biomarkers at 35–37 weeks' gestation. *Am J Obstet Gynecol* 2019; **220**: 486.e1–11.
8. Royal College of Obstetricians and Gynaecologists. The Investigation and Management of the Small-for-Gestational-Age Fetus. Green-top guideline No. 31. RCOG, 2014.
9. Poon LC, Karagiannis G, Staboulidou I, Shafiei A, Nicolaides KH. Reference range of birth weight with gestation and first-trimester prediction of small-for-gestation neonates. *Prenat Diagn* 2011; **31**: 58–65.
10. 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**: 148–154.
11. Papastefanou I, Souka AP, Pilalis A, Eleftheriades M, Michalitsi V, Kassanos D. First trimester prediction of small and large for gestation neonates by an integrated model incorporating ultrasound parameters, biochemical indices and maternal characteristics. *Acta Obstet Gynecol Scand* 2012; **91**: 104–111.
12. Crovetto F, Triunfo S, Crispi F, Rodriguez-Sureda V, Dominguez C, Figueras F, Gratacos E. Differential performance of first-trimester screening in predicting small-for-gestational-age neonate or fetal growth restriction. *Ultrasound Obstet Gynecol* 2017; **49**: 349–356.
13. Wright D, Syngelaki A, Akolekar R, Poon LC, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal characteristics and medical history. *Am J Obstet Gynecol* 2015; **213**: 62.e1–10.
14. O'Gorman N, Wright D, Syngelaki A, Akolekar R, Wright A, Poon LC, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 11–13 weeks' gestation. *Am J Obstet Gynecol* 2016; **214**: 103.e1–12.
15. Wright D, Tan MY, O'Gorman N, Poon LC, Syngelaki A, Wright A, Nicolaides KH. Predictive performance of the competing risk model in screening for preeclampsia. *Am J Obstet Gynecol* 2019; **220**: 199.e1–13.
16. Tan MY, Syngelaki A, Poon LC, Rolnik DL, O'Gorman N, Delgado JL, Akolekar R, Konstantinidou L, Tsavdaridou M, Galeva S, Ajdacka U, Molina FS, Persico N, Jani JC, Plasencia W, Greco E, Papaioannou G, Wright A, Wright D, Nicolaides KH. Screening for pre-eclampsia by maternal factors and biomarkers at 11–13 weeks' gestation. *Ultrasound Obstet Gynecol* 2018; **52**: 186–195.
17. 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**: 196–205.
18. Nicolaides KH. Screening for fetal aneuploidies at 11 to 13 weeks. *Prenat Diagn* 2011; **31**: 7–15.
19. Robinson HP, Fleming JE. A critical evaluation of sonar crown rump length measurements. *Br J Obstet Gynaecol* 1975; **82**: 702–710.
20. Nicolaides KH, Wright D, Syngelaki A, Wright A, Akolekar R. Fetal Medicine Foundation fetal and neonatal population weight charts. *Ultrasound Obstet Gynecol* 2018; **52**: 44–51.
21. Gilks WR, Thomas A, Spiegelhalter DJ. A language and program for complex Bayesian modelling. *The Statistician* 1994; **43**: 169–177.
22. R Development Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.

