

RESEARCH

Open Access



AI-based analysis of fetal growth restriction in a prospective obstetric cohort quantifies compound risks for perinatal morbidity and mortality and identifies previously unrecognized high risk clinical scenarios

Raquel M. Zimmerman¹, Edgar J. Hernandez², Mark Yandell², Martin Tristani-Firouzi³, Robert M. Silver⁴, William Grobman⁵, David Haas⁶, George Saade⁷, Jonathan Steller⁸ and Nathan R. Blue^{4*}

Abstract

Background Fetal growth restriction (FGR) is a leading risk factor for stillbirth, yet the diagnosis of FGR confers considerable prognostic uncertainty, as most infants with FGR do not experience any morbidity. Our objective was to use data from a large, deeply phenotyped observational obstetric cohort to develop a probabilistic graphical model (PGM), a type of “explainable artificial intelligence (AI)”, as a potential framework to better understand how interrelated variables contribute to perinatal morbidity risk in FGR.

Methods Using data from 9,558 pregnancies delivered at ≥ 20 weeks with available outcome data, we derived and validated a PGM using randomly selected sub-cohorts of 80% ($n = 7645$) and 20% ($n = 1,912$), respectively, to discriminate cases of FGR resulting in composite perinatal morbidity from those that did not. We also sought to identify context-specific risk relationships among inter-related variables in FGR. Performance was assessed as area under the receiver-operating characteristics curve (AUC).

Results Feature selection identified the 16 most informative variables, which yielded a PGM with good overall performance in the validation cohort (AUC 0.83, 95% CI 0.79–0.87), including among “N of 1” unique scenarios (AUC 0.81, 0.72–0.90). Using the PGM, we identified FGR scenarios with a risk of perinatal morbidity no different from that of the cohort background (e.g. female fetus, estimated fetal weight (EFW) 3–9th percentile, no preexisting diabetes, no progesterone use; RR 0.9, 95% CI 0.7–1.1) alongside others that conferred a nearly 10-fold higher risk (female fetus, EFW 3–9th percentile, maternal preexisting diabetes, progesterone use; RR 9.8, 7.5–11.6). This led to the recognition of a PGM-identified latent interaction of fetal sex with preexisting diabetes, wherein the typical protective effect of female fetal sex was reversed in the presence of maternal diabetes.

*Correspondence:

Nathan R. Blue
nblue1297@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Conclusions PGMs are able to capture and quantify context-specific risk relationships in FGR and identify latent variable interactions that are associated with large differences in risk. FGR scenarios that are separated by nearly 10-fold perinatal morbidity risk would be managed similarly under current FGR clinical guidelines, highlighting the need for more precise approaches to risk estimation in FGR.

Keywords Explainable artificial intelligence, Pregnancy, Fetal growth restriction, Stillbirth, Perinatal morbidity

Background

Fetal growth restriction (FGR) is commonly defined as estimated fetal weight (EFW) < 10th percentile and is a leading risk factor for stillbirth [1]. While FGR is an important indicator of stillbirth risk, most fetuses diagnosed with FGR do not experience any morbidity, such that the diagnosis of FGR confers a wide range of possible perinatal outcomes, from wellness to severe illness or death [2–4]. For both patients and clinicians, the high degree of uncertainty inherent in a prenatal diagnosis of FGR poses difficulties for care planning, underscoring the crucial importance of effective risk stratification for quality perinatal care. Although multivariable modeling has been studied as a potential means to improve prediction of adverse outcomes in FGR, significant progress has not been made in accurately identifying risk strata [5, 6].

Artificial intelligence (AI)-based technologies hold promise to improve estimation of perinatal morbidity risk in FGR. However, concerns exist about the transparency of AI analyses [7–9]. Reverting to human risk estimation does not resolve these concerns, as human cognition is also opaque, state-dependent, and prone to bias [10–12].

“Explainable AI” describes a subset of AI methodologies designed to improve trustworthiness by providing interpretable, transparent explanations for model decisions and predictions, thus allowing for human intellectual oversight and critical evaluation of outputs [13]. Probabilistic graphical models (PGMs) are a type of explainable AI that are well-suited to risk stratification because of their ability to capture and quantify conditional dependencies between interrelated variables in a transparent manner [14, 15]. Our objectives were (1) to derive a PGM for quantification of context-specific probabilities of composite perinatal morbidity and mortality in FGR and (2) to use the PGM to identify clinical contexts, or combinations of variables that are associated with increased or reduced morbidity risk in FGR.

Methods

This study uses data from the NICHD-funded Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b) observational cohort [16]. The nuMoM2b dataset consists of 10,038 nulliparous participants with ultrasound-confirmed, viable gestations who were recruited between 6 weeks + 0 days and 13 weeks + 6 days’ gestation. Recruitment took place at eight geographically diverse U.S. centers between October

2010 and September 2013. Participants were eligible for inclusion if they had no prior pregnancies reaching at least 20 weeks’ gestation. Exclusion criteria included age < 13 years, history of ≥ 3 pregnancy losses, the presence of a likely fatal fetal malformation already evident at enrollment screening, known fetal aneuploidy, conception using a donor oocyte, multifetal reduction, plan to terminate pregnancy, or participation in an intervention study. Participants were followed longitudinally, undergoing four study visits at the following gestational ages: Visit 1, 6w0d to 13w6d; Visit 2, 16w0d to 21w6d; Visit 3, 22w0d to 29w6d; Visit 4, after delivery. Study visits involved detailed interviews, questionnaires, research ultrasounds, maternal biometric measurements, and biospecimen collection. Ultrasounds were performed at all three study visits during pregnancy and included experimental as well as standard clinical assessments. Ultrasound estimated fetal weight (EFW) was calculated using the Hadlock four parameter formula [17], and EFW percentiles were calculated using Hadlock’s growth curve [18]. Details of medication use, medical history, and family medical history were ascertained during interviews.

In addition to clinical parameters, data collection included validated instruments that assessed multiple psychosocial domains, such as stress, social support, shift-work timing, pregnancy intendedness, and experiences of discrimination. After delivery, pre-specified obstetric, maternal, and neonatal outcomes were ascertained through maternal interviews and medical record abstraction by centrally trained research personnel. Methods of the nuMoM2b study and data collection have been described in detail [19].

All nuMoM2b participants who delivered ≥ 20 weeks’ gestation with available birth and neonatal outcomes data were included in this analysis. The only exclusion criteria were delivery prior to 20 weeks and missing birth and neonatal outcome data. The primary clinical outcome for this analysis was a composite of perinatal morbidity and mortality that was selected to reflect the consequences of fetal or neonatal compromise that sonographically identified FGR is supposed to identify. Composite morbidity was defined by the presence of any of the following factors: stillbirth, neonatal death, the need for mechanical ventilation, respiratory distress syndrome (RDS), necrotizing enterocolitis (NEC), confirmed sepsis, grade 3 or 4 intraventricular hemorrhage (IVH), neonatal seizures, or NICU admission greater than seven days. These factors

were selected from the FGR core outcome set outlined in the COSGROVE study [20]. We used a random subset of 80% of the 10,038 nuMoM2b participants for model derivation. The remaining 20% were used for model validation.

Feature selection

The initial set of potentially eligible features included >4000 unique variables. After expert review, these were pruned to a set of 907 candidate variables, which included variables from multiple domains: clinical (e.g., blood pressure, BMI, fetal sex, maternal hospitalizations during pregnancy, medication use, mode of delivery); ultrasound (visit 3 EFW percentile, visit 2 uterine artery pulsatility index, visit 2 cervical length); psychosocial (poverty, education level, experiences of discrimination, Perceived Stress scale, Edinburgh Postnatal Depression scale, State-Trait Anxiety Index-Trait scale, Multidimensional Scale of Perceived Social Support); and multi-substance quantitative urine toxicology performed for research purposes [21–25]. Missingness was not used to exclude variables, as PGMs inherently handle missing data by modeling the “unknown” state for each analyzed variable.

When possible, continuous variables were dichotomized using established clinically relevant definitions (e.g. blood pressure, cervical length). Continuous variables for which values may be associated with adverse outcomes at both upper and lower extremes (e.g. maternal age, BMI) were collapsed into multiple categories using established definitions, which were then one-hot encoded such that each category was tested as a separate dichotomous variable (present/absent). Some variables, such as hypertensive disorders of pregnancy, were tested both as dichotomized and as one-hot encoded sub-categories (gestational hypertension, mild preeclampsia, severe preeclampsia, HELLP syndrome, eclampsia, unspecified) and were selected for further testing based on the strength of their association. Continuous variables without established thresholds were assessed for association with the outcome using logistic regression, and empirical thresholds were established using receiver operating characteristics (ROC) methodology. Ultimately, fetal growth was categorized as EFW < 3rd percentile, EFW 3–9th percentile, EFW 10–90th percentile, or EFW > 90th percentile by study ultrasound done at visit 3 (22–29w6d) and using Hadlock EFW percentiles. Gestational age at birth was categorized by standard definitions of early preterm birth (PTB, < 34 weeks) late PTB (34–36 weeks), and term birth (≥ 37 weeks). Hypertensive disorders of pregnancy (HDP) were empirically dichotomized based on association with perinatal morbidity, such that the dichotomized “HDP” variable was ultimately defined as either having eclampsia,

preeclampsia with severe features, or gestational hypertension with onset prior to labor.

We chose to include all possible variables that would explain variation in perinatal morbidity risk, including those that could not be known until the time of birth, because there currently is no gold standard for FGR against which to benchmark prediction models. The adverse outcomes caused by FGR (stillbirth, perinatal morbidity, preterm birth, etc.) often occur via multiple other pathologies, and the lack of a diagnostic gold standard for FGR means it is unclear what proportion of such adverse perinatal outcomes are directly attributable to FGR. As a result, even an excellent FGR-focused model is unlikely to achieve good overall performance if it does not account for alternate causes of perinatal morbidity. This is likely the reason why studies comparing various diagnostic strategies for FGR, including FGR-specific multivariable models, have yielded statistically significant but clinically insignificant improvements in overall morbidity risk stratification performance [26–32], a finding our group has replicated repeatedly [2, 4, 5, 33, 34]. Therefore, we included all variables that might account for perinatal morbidity risk in order to better capture and understand the landscape of factors that drive risk in FGR.

Candidate features for a logistic regression (LR) model were identified with chi-square analysis and mutual information criterion (MIC) analysis. MIC quantifies the information shared between possible predictors and the target variable, and chi-square assesses independence between the possible predictor and target variable. We used a Jaccard similarity matrix to reduce redundancy, favoring variables with standard clinical definitions, when possible, to maximize clinical interpretability. The 30 variables demonstrating the strongest association with the composite primary outcome by chi-square and MIC were selected. Practical computational constraints limit exact PGMs to less than 20 variables because the number of possible combinations with additional variables expands super-exponentially (faster than any exponential function), thus becoming too computationally expensive above ~ 20 variables. To accommodate this practical constraint, we then created all possible combinations of 12 variables from the top 30 and derived a logistic regression model for each. We chose the set of 12 variables with the best-performing LR model based on the receiver-operating characteristics area under the curve (AUC).

The significant features identified by the best-performing regression model, with four additional clinical variables of interest (HDP, term birth, visit 3 EFW < 3rd percentile, and visit 3 EFW 3–9th percentile), were selected as features for the PGM model. The PGM structure was learned with the R package *bnlearn* [35], which provides a Bayesian Information Criterion-based

structure search algorithm for the creation of Bayesian networks [36–38]. The search algorithm explores the entire applicable space of conditional dependencies to discover the optimal network structure for the data. Parameter learning for this optimal network was accomplished using the Bayesian parameter estimation [39, 40]. The PGM graphical structure was also determined, allowing for visualization and greater exploration of risk factor dependencies and the impact of multiple comorbidities on the outcome of composite morbidity [35, 41]. Variables within the PGM that are mutually exclusive (e.g. categories for gestational age at birth) were blacklisted so that, for example, the absence of term birth could not be used to predict preterm birth. Our group has previously published a more technical and in-depth explanation of methods for PGM derivation [42, 43]. Interactions were identified whenever a variable's direction or magnitude

varied according to the presence of another variable to a degree that the 95% confidence intervals did not overlap. Finally, the PGM was used to identify the “Markov blanket”, or the set of variables beyond which other variables do not provide additional information about the probability of perinatal morbidity if all variables in the Markov blanket are known [44].

We used stratified 5-fold cross-validation (CV) within the 80% derivation cohort to assess performance for both LR and PGM analysis. The dataset was divided into five folds while maintaining the distribution of classes. Then, four folds of the data were used for training and the last fold was used for evaluation. This process was repeated five times. The metric used for evaluating the models' performance was the AUC. Final validation consisted of applying the PGM to the validation cohort and assessing its performance and optimal prediction threshold using ROC methodology. The exact PGM was used to estimate the absolute risk (AR) and relative risk (RR) of composite perinatal morbidity across a range of scenarios. The distribution of model risk estimation was obtained using randomly drawn samples of the data with replacement to create 1000 bootstraps. Model estimates were aggregated for all bootstraps to determine their distribution and estimate 95% confidence intervals (CI's).

Our institutional review board designated this analysis as exempt from oversight based on the definition of human subjects research. Participants gave written informed consent for the IRB approved parent study.

Results

Of 10,038 potentially eligible participants, we excluded 394 for missing neonatal outcome data and 86 for delivery prior to 20 weeks' gestation, leaving 9,558 participants who were included in the analysis. Composite perinatal morbidity occurred in 8.2% ($n=783$). Table 1 summarizes the demographic and clinical characteristics of analyzed nuMoM2b participants.

Based on feature selection from the 80% derivation cohort ($n=7,645$), the final list of variables in the PGM included the following (Table 2): maternal variables: progesterone use, pre-existing diabetes (type 1 or 2), and BP > 140/90 at visit 2 (16–21 weeks); obstetric variables: gestational age at birth (term, late preterm, early preterm), HDP (as defined above), preterm premature rupture of membranes (PPROM), urgent cesarean; and fetal/neonatal variables: sex, presence of any congenital anomaly, 5-minute Apgar < 7, and EFW percentile (< 3rd, 3-9th, 10-90th, > 90th) at visit 3 (22–29 weeks).

The PGM graphical structure is shown in Fig. 1. PGM variables had low missingness, with all variables missing less than 2% of data points except for visit 3 EFW (6.4%) and visit 2 BP > 140/90 (4.9%, Supplemental Table). Variables with direct connections to perinatal morbidity

Table 1 Cohort characteristics

Variable	Mean \pm SD or % (n)
Maternal age (years)	27.0 \pm 5.6
Race	
White	60.3 (5763)
Hispanic	16.8 (1604)
Black	13.8 (1321)
Asian	4.0 (385)
Other	5.0 (481)
Household income*	
< 100% of Fed poverty rate	12.8% (1225)
100–200% Fed poverty rate	11.7% (1114)
> 200% Fed poverty rate	56.8% (5426)
BMI	26.3 \pm 6.3
Preexisting diabetes	1.6% (151)
Hypertensive disorders	
Chronic HTN	2.5% (242)
Gestational hypertension	14.2% (1355)
PE without severe features	4.0% (385)
PE with severe features	4.1% (393)
Superimposed PE	0.5% (51)
Eclampsia	0.05% (5)
GA at delivery	
Term	90.5% (8653)
Late preterm	5.8% (556)
Early preterm	2.8% (265)
Female sex	50.9% (4868)
Congenital malformation	5.9% (561)
Composite perinatal morbidity	8.2% (783)
Stillbirth	0.6% (54)
Neonatal death	0.3% (24)
NICU stay > 7 days	4.6% (444)
RDS	3.2% (303)
Mechanical ventilation	2.7% (255)
Confirmed sepsis	0.4% (36)
ROP	0.4% (36)
Seizures	0.2% (19)
NEC	0.1% (12)
Grade 3–4 IVH	0.1% (9)

*Out of $n=7765$ with household income available

Table 2 Summary of variables included in the PGM

Variable	% (n)
Preexisting diabetes	1.6% (151)
Progesterone use	2.6% (246)
Female sex	50.9% (4868)
Congenital malformation	5.9% (561)
Visit 2 BP > 140/90	0.8% (81)
Visit 3 EFW percentile	
<3rd	1.0% (97)
3-9th	4.3% (408)
10-90th	81.0% (7745)
>90th	7.3% (696)
PPROM	2.6% (246)
Hypertensive disorders of pregnancy	10.0% (952)
GA at delivery	
Term	90.5% (8653)
Late preterm	5.8% (556)
Early preterm	2.8% (265)
Urgent cesarean	9.1% (871)
5 min Apgar <7	4.0% (382)

included early preterm birth, late preterm birth, term birth, urgent cesarean delivery, congenital anomaly, and 5-minute Apgar <7. Of note, only term birth conferred a decreased risk of the outcome, while all other directly connected variables conferred an increased risk of composite perinatal morbidity (Fig. 1A). The PGM variables in the Markov blanket for perinatal morbidity, or the set of variables beyond which other variables are not informative if all variables in the Markov blanket are known, are shown in Fig. 1B. The PGM-estimated risk relationships for each variable with composite perinatal morbidity are shown in Fig. 2 and were similar when assessed in the setting of FGR (Supplementary Fig. 1).

Model validation and applications

When applied to the validation cohort ($n=1,912$), the PGM had good performance to identify those at risk for composite perinatal morbidity (AUC 0.83, 95% CI 0.79–0.87), which was similar to the logistic regression model (AUC 0.82, 0.68–0.92, $p=0.8$ vs the PGM, Fig. 3). Both were similar to the performance estimates from 5-fold cross-validation within the derivation cohort (PGM AUC 0.83, 0.82–0.84; LR AUC 0.82, 0.81–0.83).

However, PGMs provide an advantage over LR approaches by evaluating the complete joint probability distribution of the graphical model, which includes all interdependencies among all variables included in the model. Leveraging this functionality, the PGM identified FGR scenarios in which the risk relationships varied from their typical pattern. In most clinical scenarios we analyzed, the EFW percentile category was associated with a common pattern of risk distribution, in which EFW <3rd percentile conferred the highest risk, EFW 3-9th conferred an elevated but lesser risk, and EFW >90th had

a risk that was similar to or slightly higher than the reference group of EFW 10-90th percentile (Fig. 4). However, in the setting of late PTB, EFW percentile had a distinct risk pattern from other scenarios, with all EFW percentile categories conferring a similar increase in risk (RR ≈ 2.0 , Fig. 4) compared to EFW 10-90th percentile.

We then used the PGM to estimate risk in several hypothetical scenarios involving EFW 3-9th percentile in combination with multiple other variables, which were sequentially added to identify specific combinations that drive risk. We used the PGM to compute the RR conferred by each scenario in comparison to the cohort background risk (8.3%), and to EFW 3-9th percentile alone. These results are shown in Fig. 5. Among these, the lowest-risk scenario (EFW 3-9th percentile, no diabetes, no progesterone use, female sex) had a perinatal morbidity risk estimate that was no different than the cohort background (RR 0.9, 0.7–1.1) but was lower than EFW 3-9th percentile alone (RR 0.7, 0.6–0.8). The highest-risk of the hypothetical scenarios (EFW 3-9th percentile, preexisting diabetes, fetal anomaly, progesterone use, female sex) had a risk estimate that was nearly 10-fold higher than the cohort background risk (RR 9.8, 7.5–11.6) and 8-fold higher than EFW 3-9th percentile alone (RR 7.8, 5.6–9.9, Fig. 5). The highest risk scenario estimate was based on PGM modeling alone as there were no participants in the derivation cohort meeting these criteria.

Fetal sex, diabetes, and perinatal morbidity

This series of PGM queries identified unexpected interactions between fetal sex, preexisting diabetes, and EFW percentile category. When assessed in isolation as well as in most other scenarios, female sex was either protective or had no association with composite morbidity (Supplementary Fig. 2). However, in the setting of preexisting diabetes, female sex was associated with greater risk of perinatal morbidity than male sex (RR 1.3, 1.1–1.5, Fig. 6). When applied to FGR, the associations also varied according to EFW percentile category: in the setting of EFW 3-9th percentile without preexisting diabetes, female sex was associated with lower perinatal morbidity risk compared to male sex (RR 0.8, 0.7–0.9). However, in the setting of EFW 3-9th percentile with preexisting diabetes, female sex conferred elevated risk compared to male sex (RR 1.6, 1.3–2.1). This diabetes-specific association of female sex with morbidity varied by EFW percentile category, constituting a three-way interaction. Female sex in the setting of preexisting diabetes had a similar association with morbidity in the setting of EFW <3rd percentile (RR 1.4, 1.1–1.7), 3-9th percentile (RR 1.6, 1.3–2.1), and EFW >90th percentile (RR 1.8, 1.4–2.5), but not in EFW 10-90th percentile (RR 1.11, 1.0–2.5). The RRs and absolute risk differences associated with female sex

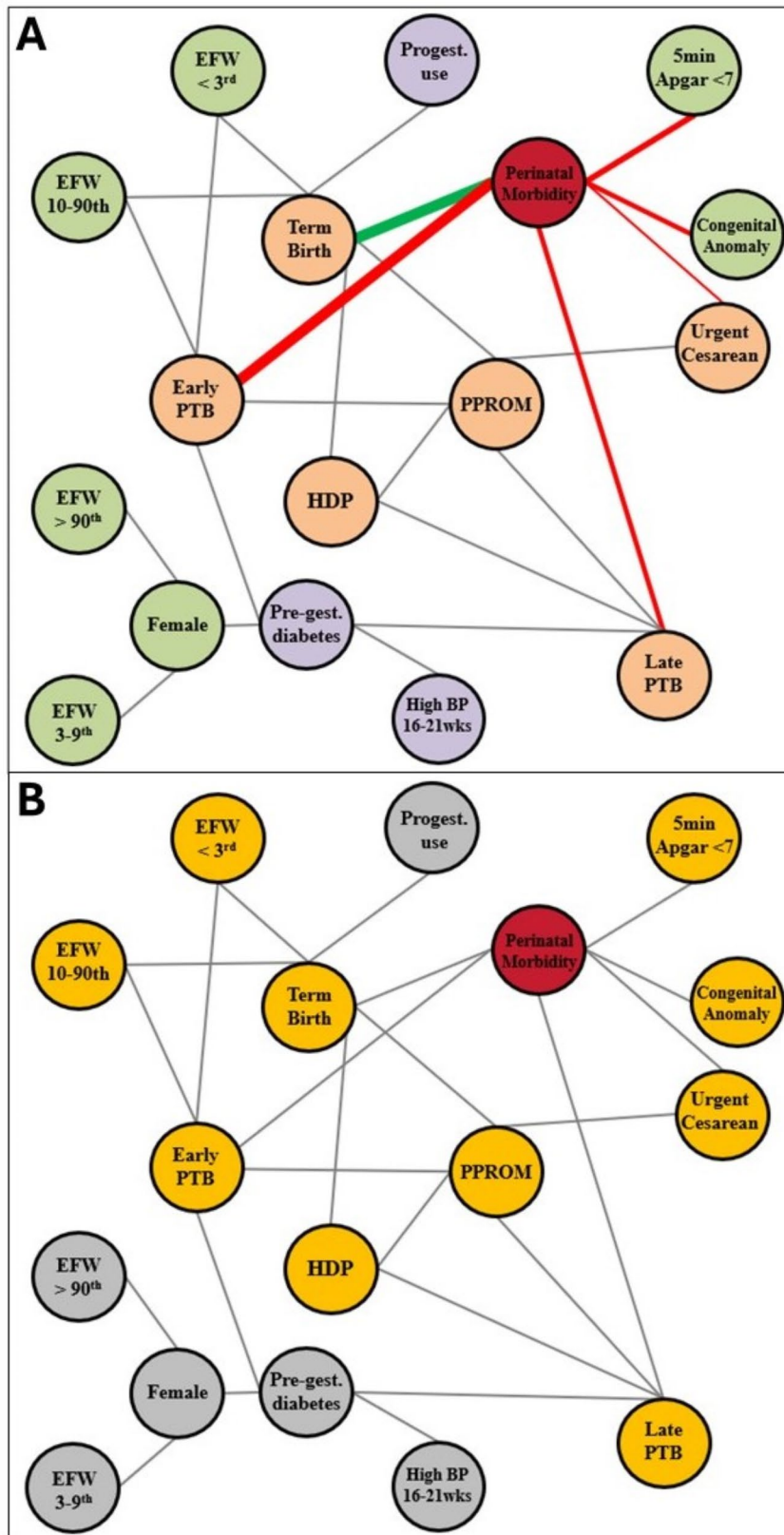


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Probabilistic graphical model structure. Panel **A**. Nodes in the model represent variables, with lines representing conditional dependencies between variables. Among variables with direct relationships to perinatal morbidity, red lines signify association with higher risk, while the green line signifies association with lower risk. The thickness of the colored lines reflects the strength of association with perinatal morbidity. Panel **B**: Variables highlighted in gold show the Markov blanket for perinatal morbidity, or the set of variables that, if all are known, fully explain the risk of perinatal morbidity. Variables outside the Markov blanket (gray) only inform the risk of perinatal morbidity when the statuses of variables in the Markov blanket are not fully known. BP: blood pressure; EFW: estimated fetal weight; HDP: hypertensive disorder of pregnancy (any of: gestational hypertension, preeclampsia, superimposed preeclampsia eclampsia); PTB: preterm birth; PPRM: preterm premature rupture of membranes

compared to male sex in FGR and preexisting diabetes are shown in Fig. 6.

PGM probability estimation for variables other than perinatal morbidity

Following our identification of the interaction between fetal sex, preexisting diabetes, and composite perinatal morbidity, we leveraged the ability of PGMs to estimate the probability of any variable in the network without having to derive a new PGM. In other words, PGMs can treat any variable within the model as either a risk factor or the target outcome for probability estimation. We used this PGM function to estimate the probability of three variables initially treated as risk factors for composite perinatal morbidity in various clinical scenarios involving FGR: early preterm birth, late preterm birth, and urgent cesarean (Fig. 7). In the setting of EFW 3-9th percentile and preexisting diabetes, the combination of progesterone use and female sex was associated with significantly increased risk for these four outcomes, compared to no progesterone use and male sex (adverse outcome RR range 2.9–3.23, Fig. 7). When compared to the cohort background, EFW 3-9th percentile alone conferred only a modest risk for these four outcomes (RR range 1.1–1.5), which was somewhat higher in the setting of preexisting diabetes and male sex without progesterone use (RR range 0.9–2.9). However, EFW 3-9th percentile in the setting of a female fetus, preexisting diabetes, and progesterone use conferred much higher risks of adverse outcomes over the cohort background (RR range 2.5–9.2). The combination of progesterone with these variables does not constitute an interaction (Supp. Figure 3) but does highlight the ability of the PGM to identify similar-appearing scenarios that are separated by large differences in risk.

N of 1 scenarios

Because several of the scenarios outlined in Figs. 4, 5, 6 and 7 included PGM queries that were represented by few or no participants in the derivation cohort, we assessed the PGM's performance in rare scenarios occurring fewer than 10 times each in the validation cohort. In the entire nuMoM2b cohort, 3.0% ($n=290$) of participants experienced a "unique scenario" consisting of a set of variables (excluding the composite morbidity outcome) not found in any other participant. In the validation cohort, there were 102 such participants (5.3%).

Among these N of 1 participants, the PGM performance remained good (AUC 0.81, 0.72–0.90), which remained the case when it was assessed among scenarios occurring fewer than or equal to 5 times (AUC 0.86, 0.81–0.91) or 10 times (0.84, 0.79–0.88).

Discussion

We developed a PGM with strong performance for composite perinatal morbidity risk estimation and the ability to estimate context-specific morbidity risk in FGR. Starting with a comprehensive assessment of clinical, psychosocial, and experimental ultrasound variables, we found that the set of variables yielding the best overall prediction largely consisted of variables that are routinely ascertained in clinical practice. When using the PGM to explore the risk landscape in FGR based on ultrasound at 22–29 weeks, we found that the PGM captured and quantified an unexpected interaction between preexisting diabetes, fetal sex, and composite perinatal morbidity. The PGM was able to identify scenarios that appear similar clinically but have large differences in morbidity risk. Finally, it maintained good performance among the 102 "N of 1" scenarios in the validation cohort.

The variables retained in the PGM for prediction of perinatal morbidity and mortality are largely consistent with existing literature. The association between progesterone use and increased risk of perinatal morbidity was not surprising since we interpret progesterone use as reflecting increased clinical concern for PTB or miscarriage. While we expected that psychosocial factors would be more informative for risk estimation, the variables retained in the PGM are those most closely associated with composite perinatal morbidity and may be on the causal pathway between social determinants of health and adverse outcomes. If a hypothetical PGM were to include many more variables, psychosocial factors may be retained. However, because PGMs have a practical limit of <20 variables owing to computational constraints, these variables were not included, reflecting a limitation of the PGM approach.

The inclusion of the EFW <3rd percentile and EFW 10-90th percentile variables within the Markov blanket indicates that they still provide unique information about the probability of morbidity, even when all other Markov blanket variables are known. However, the absence of direct connections between the EFW 3-9th and EFW >90th percentiles and perinatal morbidity suggests that

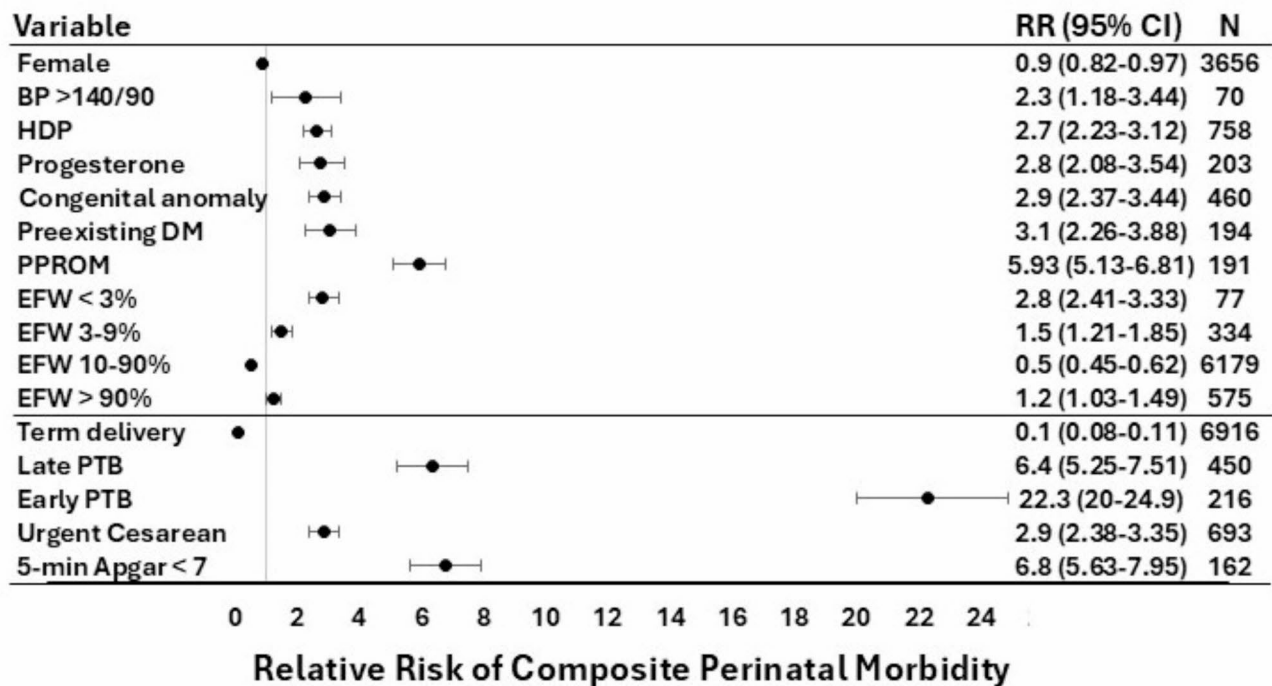


Fig. 2 PGM-estimated composite perinatal morbidity risk conferred by individual PGM variables. The vertical gray line reflects the reference (RR=1), based on the cohort's background risk of composite perinatal morbidity. BP > 140/90 was ascertained at visit 2 (16–21 weeks). Factors known prior to delivery (above the horizontal line) are separated from factors only known at or after delivery (below the line). In the EFW categories, “%” denotes percentile. BP: blood pressure; CI: confidence interval; DM: diabetes mellitus; EFW: estimated fetal weight; HDP: hypertensive disorder of pregnancy (any of: gestational hypertension, preeclampsia, superimposed preeclampsia, eclampsia); PGM: probabilistic graphical model; PTB: preterm birth; PPRM: preterm premature rupture of membranes; RR, relative risk

their relationships to morbidity are mediated by other variables within the PGM. Finally, our finding of a diabetes-specific relationship between fetal sex and perinatal morbidity was striking and unexpected. This finding is likely robust given that it was based on a relatively large number of participants (Fig. 6, $n=3598$ for female sex without diabetes, $n=58$ for female sex with diabetes). In the PGM, this interaction appeared to vary by EFW percentile category, suggesting a 3-way interaction. However, the number of participants meeting those query criteria was low or zero, such that the estimates suggesting a 3-way interaction may be less trustworthy. There is an emerging body of literature describing sex-specific responses to maternal metabolic perturbations, including hyperglycemia and obesity, placental gene expression [45], fatty acid metabolism [46], protein metabolism [47], and fetal growth [48, 49]. Findings on the association of fetal sex with perinatal morbidity in gestational diabetes are mixed [48–50], and it is uncertain whether the interaction we describe would be expected to generalize to the context of gestational diabetes, which is the focus of most available studies.

To date, investigations of explainable AI in reproductive science have been limited. Broader machine learning (ML) approaches have been used for stillbirth, FGR, and

preeclampsia prediction as well as basic science applications such as multi-omic placental phenotyping, among others [51–56]. There have been other efforts to utilize explainable ML approaches for prediction of gestational diabetes [57], preterm birth [57], and composite perinatal morbidity [58], though these efforts consisted of developing predictive ML models using traditional approaches, followed by post-hoc application of a secondary approach to identify the most contributory variables and thereby explain the model's prediction. For example, applying Shapley values to a model improves its explainability by quantifying each feature's individual contribution to the model's output (“feature importance”), but does not provide clarity on the interdependent relationships between exposure variables or identify scenarios characterized by synergistic or non-linear changes in probability. In contrast, PGMs handle this explainability inherently. The explainability of PGMs is apparent in that a given variable's precise and context-specific statistical contribution to risk is transparently quantified and reported when risk is estimated both with and without the variable of interest. The use of PGMs in reproductive medicine thus far has been limited to optimizing assisted reproductive technologies [59] and prediction of neonatal pneumonia in the setting of maternal diabetes [60].

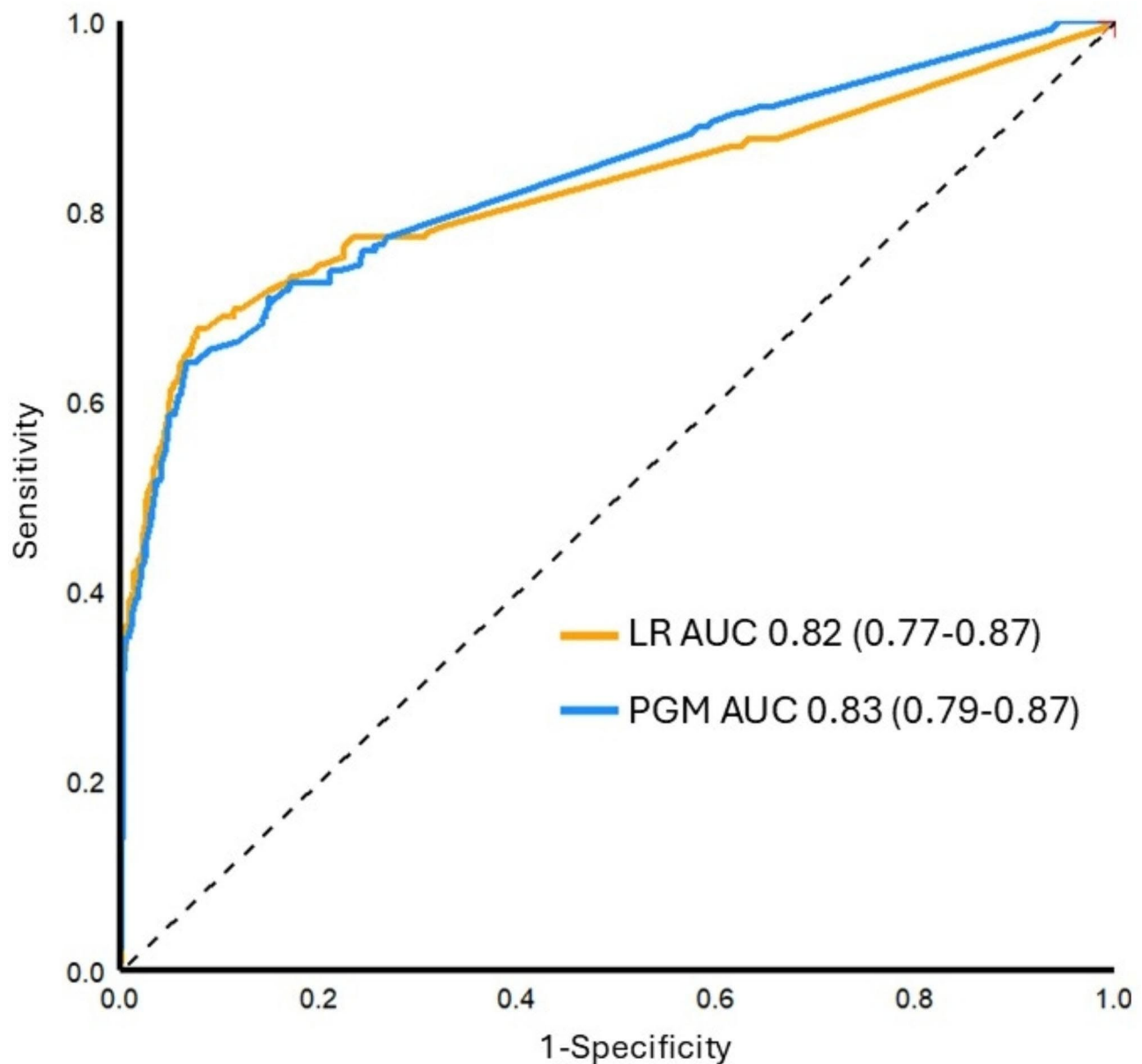


Fig. 3 Receiver operating characteristics curves for the PGM and logistic regression model and for prediction of composite perinatal morbidity in the validation cohort. ROC: receiver operating characteristics; PGM: probabilistic graphical model; LR: logistic regression; AUC: area under the curve. AUC 95% confidence intervals shown in parentheses

A key advantage of PGMs is their ability to produce relatively precise probability estimates for low frequency or “N of 1” scenarios. Because PGMs can capture and quantify conditional dependencies between interrelated variables, the estimated influence of a given variable is adjusted based on the presence or absence of other factors linked to the given variable via such conditional dependencies. This allows for reproducible and relatively precise probability estimations in the types of “N of 1” scenarios that individually are uncommon but collectively are common in clinical practice. Currently, risk stratification in such scenarios depends on human expert

assessment, which is highly flexible but opaque, prone to bias, and influenced by circumstances such as sleep deprivation, mood, and other factors [10–12]. In contrast, PGMs overcome the opaque and state-dependent pitfalls of human risk estimation thanks to their transparency and reproducibility. PGMs can therefore be used to transparently generate probability estimates, including for rarely occurring scenarios, while human experts can provide both intellectual oversight and patient-centered recommendations.

The U.S. clinical guidelines for management of FGR include the frequency of monitoring and timing of

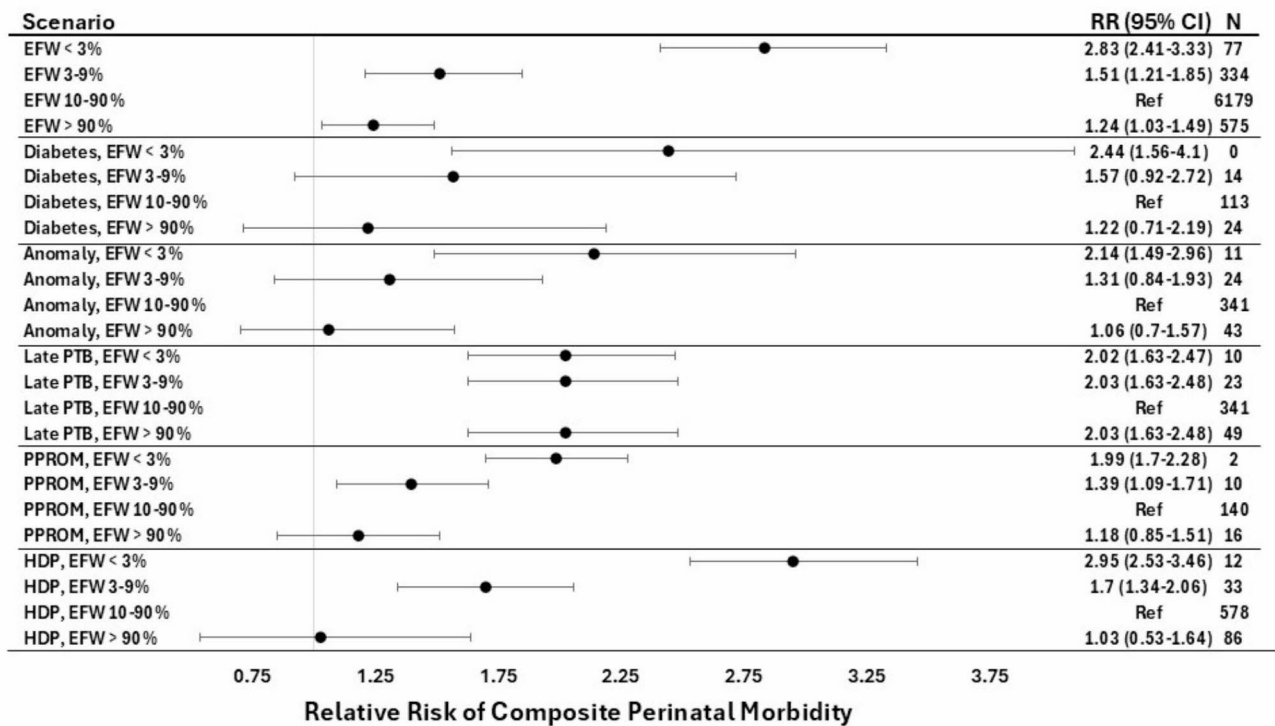


Fig. 4 Relative risk of perinatal morbidity conferred by EFW percentile category across a range of obstetric scenarios. The RR for each high or low EFW percentile category was compared against EFW 10th -90th percentile in the setting of the same clinical scenario, labeled as “Ref.” In the clinical scenarios, “%” denotes percentile. The vertical gray line reflects RR of 1. N’s report the number of participants in the derivation cohort with the associated clinical scenario. Diabetes refers to pre-gestational diabetes. Point estimates are based on the PGM’s maximum likelihood estimates rather than the mean of bootstrapped values, which is why they are not in the center of the confidence intervals. CI: confidence interval; EFW: estimated fetal weight; HDP: hypertensive disorder of pregnancy (any of: gestational hypertension: preeclampsia: superimposed preeclampsia: eclampsia); PTB: preterm birth; PPRM: preterm premature rupture of membranes; RR: relative risk

delivery and are based on metrics of FGR severity alone. These guidelines do not take other clinical data into account, allowing clinicians to customize their recommendation using clinical judgement [61, 62]. We identified several clinical scenarios, all involving non-severe FGR (EFW 3-9th percentile) that were separated by up to a 10-fold difference in perinatal morbidity risk (Fig. 5) for which the U.S. guidelines recommend identical management. This reality highlights current gaps and opportunities for personalized risk estimation to reduce prognostic uncertainty of FGR and tailor surveillance and delivery timing plans accordingly.

Our study’s strengths included the comprehensive assessment of variables across domains, including clinical, SDoH, mental health, ultrasound, and quantitative urine toxicology in a large and deeply phenotyped pregnancy cohort. We used a novel explainable AI approach to produce a flexible model with good prediction of composite perinatal morbidity and the potential for integration of genomic and other variables. The cohort is geographically diverse in the U.S. and utilized standardized outcomes ascertainment by centrally trained and certified research personnel.

Limitations to generalizability include the timing of research ultrasounds at 22–29 weeks and the lack of other variables informative for risk, such as umbilical artery Doppler assessments and maternal biomarkers. Additionally, nuMoM2b, while diverse, is not fully representative of the U.S. population. Given that EFW percentile estimates used in our study were collected for investigational purposes only and were not routinely disclosed to clinicians or participants [19], we cannot rule out that clinical recognition and management of FGR may have introduced bias that affects our analysis of the associations between EFW percentile and perinatal morbidity. Progesterone use is an intervention based almost solely on clinical concern and thus likely includes bias, making its associations with morbidity difficult to interpret. We included it because our goal was to derive a PGM using the most empirically informative set of factors for morbidity risk estimation, but the association should be interpreted as hypothesis-generating only and not potentially causal. Also, the fact that the progesterone use variable was outside the Markov blanket (Fig. 1B) means that it only informs perinatal morbidity risk estimates when other Markov blanket variables, such as gestational age at

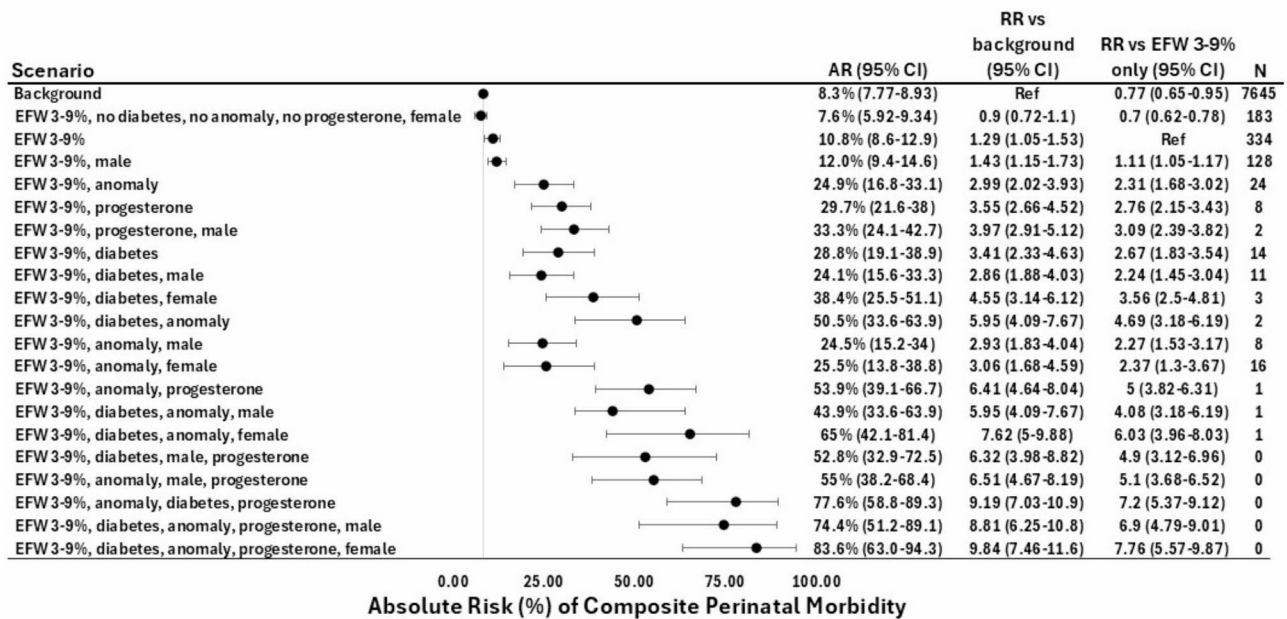


Fig. 5 Sequential introduction of clinical factors to non-severe FGR to identify variable combinations that drive composite perinatal morbidity risk. RR columns are based on a given scenario's comparison to the cohort's background or to the risk of EFW 3-9th percentile alone (in red). The vertical gray line reflects the cohort's background risk of composite perinatal morbidity (8.3%). In the clinical scenarios on the left, "%" denotes percentile. Point estimates are based on the PGM's maximum likelihood estimates rather than the mean of bootstrapped values, which is why they are not in the center of the confidence intervals. Diabetes refers to pre-gestational diabetes. N values represent the number of participants in the derivation cohort who meet the query criteria. AR: absolute risk (expressed as a percent); CI: confidence interval; RR: relative risk; EFW: estimated fetal weight; PTB: preterm birth; PPRM: preterm premature rupture of membranes

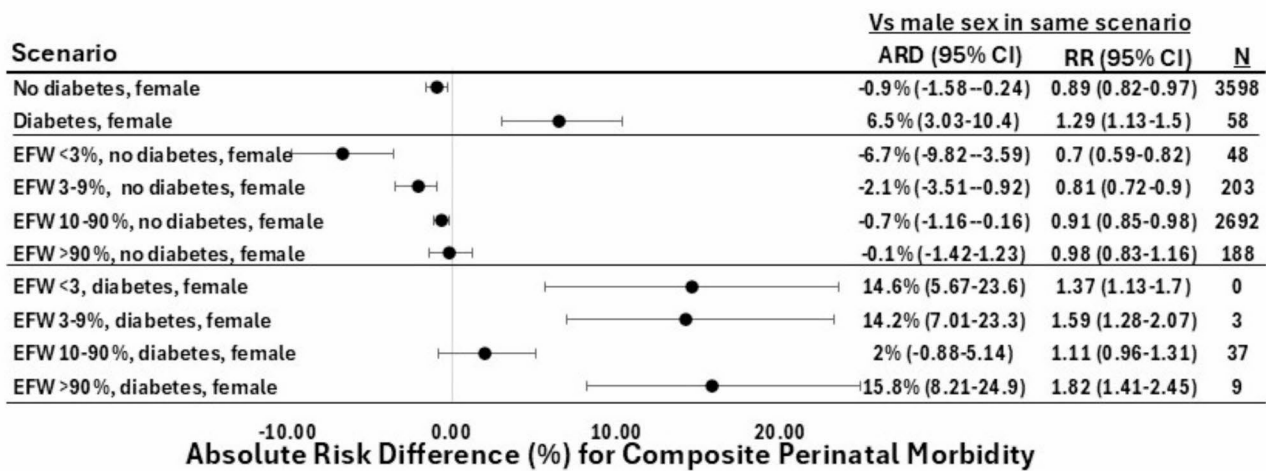


Fig. 6 Female sex confers lower perinatal morbidity risk except in the setting of pre-gestational diabetes. In the absence of pregestational diabetes, female sex is protective. In the presence of pregestational diabetes, female sex adds risk. ARD reflects the absolute risk difference (expressed as a percent) for perinatal composite morbidity between female and male sex in the given EFW percentile and diabetes scenarios. The vertical gray line reflects the risk associated with male sex in the same clinical scenario. Point estimates are based on the PGM's maximum likelihood estimates rather than the mean of bootstrapped values, which is why they are not in the center of the confidence intervals. N values represent the number of participants in the derivation cohort who meet the query criteria. Diabetes refers to pre-gestational diabetes. ARD: absolute risk difference; CI: confidence interval; RR: relative risk; EFW: estimated fetal weight

birth, are unknown. This supports our interpretation that it reflects clinical concern rather than having any causal contribution to morbidity. Finally, our inclusion of variables known only at or after delivery means this model

is not useful prenatally. We chose this approach because our objective was not to develop a tool for prenatal use to alter clinical management, but to determine the utility of PGMs to capture and quantify complex risk relationships

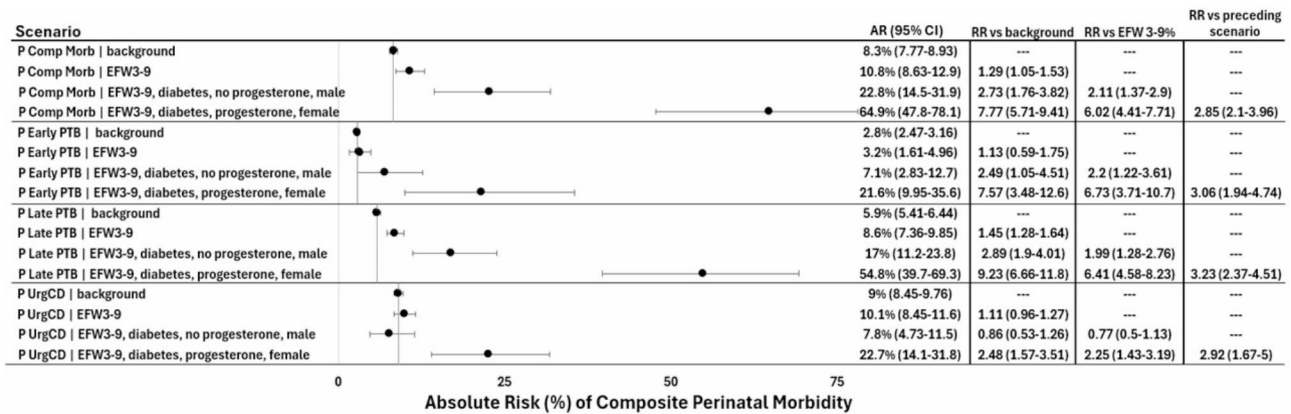


Fig. 7 PGM risk estimates for multiple adverse outcomes in FGR in the setting of pre-gestational diabetes vary widely according to progesterone use and fetal sex. The scenarios are framed as probability expressions where “P Comp Morb | EFW 3–9” would be written as “the probability of composite morbidity given the presence of EFW 3–9th percentile.” The green lines represent the cohort’s background absolute risk of the associated outcome (composite perinatal morbidity, early PTB, late PTB, or urgent cesarean), allowing for visual interpretation based on confidence intervals that overlap with the background risk estimate. Absolute risks are expressed as percentages. The number of derivation cohort participants meeting criteria for each scenario in order of presentation are: 7645, 334, 11, and 0, respectively. “RR vs. background” expresses the relative risk of a given factor or scenario over the cohort’s background risk of the same outcome (green line). “RR vs. EFW 3–9%” expresses the relative risk of each scenario over the risk conferred by EFW 3–9th percentile alone. “RR vs. preceding scenario” expresses the relative risk of the final scenario over the preceding scenario, in which the only differing factors are progesterone use and fetal sex. All risks (AR, RR) are followed by 95% confidence intervals. Diabetes refers to pre-gestational diabetes. AR: absolute risk; CI: confidence interval; RR: relative risk; EFW: estimated fetal weight; P: probability; PTB: preterm birth; UrgCD: urgent cesarean delivery

and identify context-specific risks in FGR. Eventually, this line of inquiry may lead to the development of a tool for prenatal use.

Conclusions

We successfully developed an explainable AI model with good performance for perinatal morbidity risk estimation and the ability to provide context-specific risk estimates across a range of FGR scenarios, including those that occur at a low frequency. While not yet ready for clinical application, this represents an important proof of concept and demonstration of the potential for PGMs to refine risk estimation in obstetrics.

Abbreviations

- AI Artificial intelligence
- AR Absolute risk
- ARD Absolute risk difference
- AUC Area under the curve
- BP Blood pressure
- CI Confidence interval
- CV Cross-validation
- DM (preexisting/pregestational) diabetes mellitus
- EFW Estimated fetal weight
- FGR Fetal growth restriction
- GA Gestational age
- HDP Hypertensive disorder of pregnancy
- HTN Hypertension
- IVH Intraventricular hemorrhage
- LR Logistic regression
- ML Machine learning
- NEC Necrotizing enterocolitis
- NICU Neonatal intensive care unit
- PE Preeclampsia
- PGM Probabilistic graphical model
- PPROM Preterm premature rupture of membranes
- PTB Preterm birth

- RDS Respiratory distress syndrome
- ROP Retinopathy of prematurity
- ROC Receiver operating characteristics
- RR Relative risk
- UrgCD Urgent cesarean delivery

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12884-024-07095-6>.

Supplementary Material 1

Author contributions

Study conception, design, analysis: RMZ, EJH, MY, MTF, NRB. Data interpretation, manuscript preparation and editing: RMZ, EJH, MY, MTF, RMS, WG, DH, GS, JS, NRB.

Funding

This work was funded by the NICHD (award numbers U10 HD063020, U10 HD063037, U10 HD063041, U10 HD063046, U10 HD063047, U10 HD063048, U10 HD063053, U10 HD063072, 2K12 HD085816-07), R Baby Foundation, and the One Utah Data Science Hub Seed Grant Program.

Data availability

NuMoM2b study data are available on request at the Data Access Specimen Hub (DASH), hosted by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, <https://dash.nichd.nih.gov>.

Declarations

Ethics approval

This study was designated by the University of Utah Institutional Review Board as exempt from oversight based on the definition of human subjects research. Participants gave written informed consent for the IRB approved parent study. Our study was carried out in compliance with the Helsinki Declaration.

Competing interests

The authors declare no competing interests.

AI use

Generative AI, including large language models, were not used in the writing of this manuscript.

Author details

¹Department of Biomedical Informatics, University of Utah Health, Salt Lake City, UT, USA

²Utah Center for Genetic Discovery, Department of Human Genetics, University of Utah Health, Salt Lake City, UT, USA

³Department of Pediatrics, Division of Pediatric Cardiology, University of Utah Health, Salt Lake City, UT, USA

⁴Department of Obstetrics and Gynecology, Division of Maternal-Fetal Medicine, University of Utah Health, 30 N. Mario Capecchi Dr., Level 5 South, Salt Lake City, UT 84132, USA

⁵Department of Obstetrics and Gynecology, Warren Alpert Medical School, Brown University, Providence, RI, USA

⁶Department of Obstetrics and Gynecology, Indiana University, Indianapolis, IN, USA

⁷Department of Obstetrics and Gynecology, Eastern Virginia Medical School, Norfolk, VA, USA

⁸Department of Obstetrics & Gynecology, Division of Maternal Fetal Medicine, University of California, Irvine, Orange, CA, USA

Received: 21 September 2024 / Accepted: 19 December 2024

Published online: 30 January 2025

References

- Flenady V, Koopmans L, Middleton P, et al. Major risk factors for stillbirth in high-income countries: a systematic review and meta-analysis. *Lancet Apr*. 2011;16(9774):1331–40. [https://doi.org/10.1016/s0140-6736\(10\)62233-7](https://doi.org/10.1016/s0140-6736(10)62233-7)
- Blue NR, Grobman WA, Larkin JC, et al. Customized versus Population Growth standards for Morbidity and Mortality Risk Stratification using Ultrasonographic fetal growth Assessment at 22 to 29 weeks' Gestation. *Am J Perinatol Aug*. 2021;38(01):e46–56. <https://doi.org/10.1055/s-0040-1705114>
- Blue NR, Beddow ME, Savabi M, Katukuri VR, Mozurkewich EL, Chao CR. A comparison of methods for the diagnosis of fetal growth restriction between the Royal College of Obstetricians and gynaecologists and the American College of Obstetricians and gynecologists. *Obstet Gynecol May*. 2018;131(5):835–41. <https://doi.org/10.1097/AOG.0000000000002564>
- Blue NR, Beddow ME, Savabi M, Katukuri VR, Chao CR. Nov. Comparing the Hadlock fetal growth standard to the Eunice Kennedy Shriver National Institute of Child Health and Human Development racial/ethnic standard for the prediction of neonatal morbidity and small for gestational age. *Am J Obstet Gynecol*. 2018;219(5):474 e1–474 e12. <https://doi.org/10.1016/j.ajog.2018.08.011>
- Blue NR, Allshouse AA, Grobman WA et al. Developing a predictive model for perinatal morbidity among small for gestational age infants. *J Matern Fetal Neonatal Med*. 2021;1–10. <https://doi.org/10.1080/14767058.2021.1980533>
- Ross C, Deruelle P, Pontvianne M, et al. Prediction of adverse neonatal adaptation in fetuses with severe fetal growth restriction after 34 weeks of gestation. *Eur J Obstet Gynecol Reproductive Biology*. 2024;296:258–64. <https://doi.org/10.1016/j.ejogrb.2024.03.008>
- Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. *J Glob Health Dec*. 2019;9(2):010318. <https://doi.org/10.7189/jogh.09.020318>
- Chen F, Wang L, Hong J, Jiang J, Zhou L. Unmasking bias in artificial intelligence: a systematic review of bias detection and mitigation strategies in electronic health record-based models. *J Am Med Inf Assoc Apr*. 2024;19(5):1172–83. <https://doi.org/10.1093/jamia/ocae060>
- Gichoya JW, Thomas K, Celi LA, et al. AI pitfalls and what not to do: mitigating bias in AI. *Br J Radiol Oct*. 2023;96(1150):20230023. <https://doi.org/10.1259/bjr.20230023>
- Dehon E, Weiss N, Jones J, Faulconer W, Hinton E, Sterling S. A systematic review of the impact of physician implicit racial bias on clinical decision making. *Acad Emerg Med Aug*. 2017;24(8):895–904. <https://doi.org/10.1111/acem.13214>
- Jala S, Fry M, Elliott R. Cognitive bias during clinical decision-making and its influence on patient outcomes in the emergency department: a scoping review. *J Clin Nurs Oct*. 2023;32(19–20):7076–85. <https://doi.org/10.1111/jocn.16845>
- Beldhuis IE, Marapin RS, Jiang YY, et al. Cognitive biases, environmental, patient and personal factors associated with critical care decision making: a scoping review. *J Crit Care Aug*. 2021;64:144–53. <https://doi.org/10.1016/j.jcrcr.2021.04.012>
- Angelov PP, Soares EA, Jiang R, Arnold NI, Atkinson PM. Explainable artificial intelligence: an analytical review. *WIREs Data Min Knowl Discov*. 2021;11(5). <https://doi.org/10.1002/widm.1424>
- Mihaljević B, Bielza C, Larrañaga P. Bayesian networks for interpretable machine learning and optimization. *Neurocomputing*. 2021;456:648–65. <https://doi.org/10.1016/j.neucom.2021.01.138>
- Ali S, Abuhmed T, El-Sappagh S, et al. Explainable Artificial Intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. *Inform Fusion*. 2023;99. <https://doi.org/10.1016/j.inffus.2023.101805>
- Haas DM, Parker CB, Wing DA, et al. A description of the methods of the nulliparous pregnancy outcomes study: monitoring mothers-to-be (nuMoM2b). *Am J Obstet Gynecol Apr*. 2015;212(4):539.e1–539.e24. <https://doi.org/10.1016/j.ajog.2015.01.019>
- Hadlock FP, Deter RL, Harrist RB, Park SK. Estimating fetal age: computer-assisted analysis of multiple fetal growth parameters. *Radiology*. 1984;152(2):497–501. <https://doi.org/10.1148/radiology.152.2.6739822>
- Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology*. 1991;181(1):129–33. <https://doi.org/10.1148/radiology.181.1.1887021>
- Haas DM, Parker CB, Wing DA et al. Apr. A description of the methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b). *Am J Obstet Gynecol*. 2015;212(4):539 e1–539 e24. <https://doi.org/10.1016/j.ajog.2015.01.019>
- Healy P, Gordijn SJ, Ganzevoort W, et al. A core outcome set for the prevention and treatment of fetal GROwth restriction: deVelooping endpoints: the COSGROVE study. *Am J Obstet Gynecol*. 2019;221(4):339e.1–339.e10. <https://doi.org/10.1016/j.ajog.2019.05.039>
- Metz TD, Allshouse AA, McMillin GA, et al. Cannabis exposure and adverse pregnancy outcomes related to placental function. *JAMA*. 2023;330(22):2191. <https://doi.org/10.1001/jama.2023.21146>
- Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Social Behav Dec*. 1983;24(4):385–96.
- Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. *Br J Psychiatry*. 1987;150(6):782–6. <https://doi.org/10.1192/bjp.150.6.782>
- Spielberger CD. Manual for the State-Trait anxiety inventory. 1983.
- Zimet GD, Powell SS, Farley GK, Werkman S, Berkoff KA. Psychometric characteristics of the multidimensional scale of perceived social support. *J Pers Assess Winter*. 1990;55(3–4):610–7. <https://doi.org/10.1080/00223891.1990.9674095>
- Gleason JL, Reddy UM, Chen Z, et al. Comparing population-based fetal growth standards in a US cohort. *Am J Obstet Gynecol Dec*. 2023;25. <https://doi.org/10.1016/j.ajog.2023.12.034>
- Monier I, Ego A, Benachi A, et al. Unisex versus sex-specific estimated fetal weight charts for fetal growth monitoring: a population-based study. *Am J Obstet Gynecol MFM*. 2021;100527. <https://doi.org/10.1016/j.ajogmf.2021.10.0527>
- Saviron-Cornudella R, Esteban LM, Tajada-Duaso M, et al. Detection of adverse perinatal outcomes at term delivery using ultrasound estimated percentile weight at 35 weeks of gestation: comparison of five fetal growth standards. *Fetal Diagn Ther*. 2020;47(2):104–14. <https://doi.org/10.1159/000500453>
- Kabiri D, Romero R, Gudicha DW, et al. Prediction of adverse perinatal outcome by fetal biometry: comparison of customized and population-based standards. *Ultrasound Obstet Gynecol*. 2020;55(2):177–88. <https://doi.org/10.1002/uog.20299>
- Zhu C, Ren YY, Wu JN, Zhou QJ. A comparison of prediction of adverse perinatal outcomes between Hadlock and INTERGROWTH-21(St) standards at the third trimester. *Biomed Res Int*. 2019;2019:7698038. <https://doi.org/10.1155/2019/7698038>
- Strassberg ER, Schuster M, Rajaram AM, et al. Comparing diagnosis of fetal growth restriction and the potential impact on management and outcomes using different growth curves. *J Ultrasound Med*. 2019;38(12):3273–81. <https://doi.org/10.1002/jum.15063>
- Rousseau T, Durand-Maison O, Labruere-Chazal C, et al. Customized and non-customized live-born birth-weight curves of single and uncomplicated pregnancies from the Burgundy perinatal network. Part I – methodology. *J Gynecol Obstet Hum Reprod*. 2017;46(7):587–90. <https://doi.org/10.1016/j.jogoh.2017.05.004>

33. Blue NR, Mele L, Grobman WA, et al. Predictive performance of newborn small for gestational age by a United States intrauterine vs birthweight-derived standard for short-term neonatal morbidity and mortality. *Am J Obstet Gynecol MFM* May. 2022;4(3):100599. <https://doi.org/10.1016/j.ajogmf.2022.100599>
34. Blue NR, Allshouse AA, Heerboth S, et al. Derivation and assessment of a sex-specific fetal growth standard. *J Maternal-Fetal Neonatal Med*. 2022:1–9. <https://doi.org/10.1080/14767058.2022.2075696>
35. Scutari M. Learning Bayesian networks with thebnlearnRPackage. *J Stat Softw* 07/16. 2010;35(3):1–22. <https://doi.org/10.18637/jss.v035.i03>
36. Yuan CMB, Wu X. Learning optimal Bayesian networks using A* search. presented at: twenty-second international joint conference on artificial intelligence. 2011.
37. Koivisto MSK. Exact Bayesian structure discovery in Bayesian networks. *J Mach Learn Res*. 2004;5:549–73. <https://doi.org/10.5555/1005332.1005352>
38. Heckerman D, Geiger D, Chickering DM. *Mach Learn*. 1995;20(3):197–243. <https://doi.org/10.1023/a:1022623210503>
39. Bretthorst GL. An introduction to parameter estimation using Bayesian probability theory. In: Fougère PF, editor. *Maximum entropy and Bayesian methods*. Springer Netherlands; 1990. pp. 53–79.
40. Scutari M. Bayesian network constraint-based structure learning algorithms: parallel and optimized implementations in the bnlearn R Package. *J Stat Softw*. 2017;77(2):1–20. <https://doi.org/10.18637/jss.v077.i02>
41. BioRender.com. 2023.
42. Watkins WS, Hernandez EJ, Wesolowski S, et al. De novo and recessive forms of congenital heart disease have distinct genetic and phenotypic landscapes. *Nat Commun* Oct. 2019;17(1):4722. <https://doi.org/10.1038/s41467-019-1258-2-y>
43. Wesolowski S, Lemmon G, Hernandez EJ, et al. An explainable artificial intelligence approach for predicting cardiovascular outcomes using electronic health records. *PLoS Digit Health*. 2022;1(1):e0000004. <https://doi.org/10.1371/journal.pdig.0000004>
44. Probabilistic Graphical Models. *Advances in computer vision and pattern recognition*. 2021.
45. Barke TL, Money KM, Du L, et al. Sex modifies placental gene expression in response to metabolic and inflammatory stress. *Placenta*. 2019;78:1–9. <https://doi.org/10.1016/j.placenta.2019.02.008>
46. Watkins OC, Yong HEJ, Mah TKL, et al. Sex-dependent regulation of placental oleic acid and palmitic acid metabolism by maternal glycemia and associations with birthweight. *Int J Mol Sci*. 2022;23(15):8685. <https://doi.org/10.3390/ijms23158685>
47. Ren ZR, Luo SS, Qin XY, Huang HF, Ding GL. Sex-specific alterations in placental proteomics induced by intrauterine hyperglycemia. *J Proteome Res* Apr. 2024;5(4):1272–84. <https://doi.org/10.1021/acs.jproteome.3c00735>
48. Gilron S, Gabbay-Benziv R, Khoury R. Same disease - different effect: maternal diabetes impact on birth weight stratified by fetal sex. *Arch Gynecol Obstet* Mar. 2024;309(3):1001–7. <https://doi.org/10.1007/s00404-023-06973-2>
49. Seghieri G, Di Cianni G, Gualdani E, De Bellis A, Franconi F, Francesconi P. The impact of fetal sex on risk factors for gestational diabetes and related adverse pregnancy outcomes. *Acta Diabetol* May. 2022;59(5):633–9. <https://doi.org/10.1007/s00592-021-01836-1>
50. Cidade-Rodrigues C, Chaves C, Melo A, et al. Association between foetal sex and adverse neonatal outcomes in women with gestational diabetes. *Arch Gynecol Obstet* Apr. 2024;309(4):1287–94. <https://doi.org/10.1007/s00404-023-06979-w>
51. Barak O, Lovelace T, Piekos S, et al. Integrated unbiased multiomics defines disease-independent placental clusters in common obstetrical syndromes. *BMC Med*. 2023;21(1). <https://doi.org/10.1186/s12916-023-03054-8>
52. Malacova E, Tippaya S, Bailey HD, et al. Stillbirth risk prediction using machine learning for a large cohort of births from Western Australia, 1980–2015. *Sci Rep*. 2020;10(1). <https://doi.org/10.1038/s41598-020-62210-9>
53. Schmidt LJ, Rieger O, Neznansky M, et al. A machine-learning-based algorithm improves prediction of preeclampsia-associated adverse outcomes. *Am J Obstet Gynecol*. 2022;227(1):77e.1-77.e30. <https://doi.org/10.1016/j.ajog.2022.01.026>
54. Sufriyana H, Amani FZ, Al Hajiri AZZ, Wu Y-W, Su EC-Y. Prognosticating fetal growth restriction and small for gestational age by medical history. *IOS; 2024*.
55. Bahado-Singh RO, Yilmaz A, Bisgin H, et al. Artificial intelligence and the analysis of multi-platform metabolomics data for the detection of intrauterine growth restriction. *PLoS ONE*. 2019;14(4):e0214121. <https://doi.org/10.1371/journal.pone.0214121>
56. Crockart IC, Brink LT, du Plessis C, Odendaal HJ. Classification of intrauterine growth restriction at 34–38 weeks gestation with machine learning models. *Inf Med Unlocked*. 2021;23. <https://doi.org/10.1016/j.imu.2021.100533>
57. Du Y, Rafferty AR, McAuliffe FM, Wei L, Mooney C. An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus. *Sci Rep*. 2022;12(1). <https://doi.org/10.1038/s41598-022-05112-2>
58. Lee SJ, Garcia GP, Stanhope KK, Platner MH, Boulet SL. Interpretable machine learning to predict adverse perinatal outcomes: examining marginal predictive value of risk factors during pregnancy. *Am J Obstet Gynecol MFM* Oct. 2023;5(10):101096. <https://doi.org/10.1016/j.ajogmf.2023.101096>
59. Hernández-González J, Valls O, Torres-Martín A, Cerquides J. Modeling three sources of uncertainty in assisted reproductive technologies with probabilistic graphical models. *Comput Biol Med* Nov. 2022;150:106160. <https://doi.org/10.1016/j.combiomed.2022.106160>
60. Lin Y, Chen JS, Zhong N, Zhang A, Pan H. A Bayesian network perspective on neonatal pneumonia in pregnant women with diabetes mellitus. *BMC Med Res Methodol*. 2023;23(1). <https://doi.org/10.1186/s12874-023-02070-9>
61. Fetal Growth Restriction. ACOG practice bulletin, number 227. *Obstet Gynecol*. Feb. 2021;1(2):e16–28. <https://doi.org/10.1097/aog.0000000000004251>
62. Society for Maternal-Fetal Medicine. Electronic address pso, Martins JG, Biggio JR, Abuhamad A. Oct. Society for Maternal-Fetal Medicine Consult Series #52: Diagnosis and management of fetal growth restriction: (Replaces Clinical Guideline Number 3, April 2012). *Am J Obstet Gynecol*. 2020;223(4):B2-B17. <https://doi.org/10.1016/j.ajog.2020.05.010>

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.