A systematic review of first trimester prediction models for gestational diabetes mellitus (GDM)

Zhiwen Liu
Supervisor- Ewoud Schuit, GP Shceffer, Ellis Becking, Jacintha van Eekhout,

OBJECTIVE: The aim of this study is to perform a systematic review of available obstetric first trimester prediction models based on maternal characteristics for gestational diabetes mellitus (GDM).
METHODS: The review included a comprehensive search of following electronic bibliographic databases: PubMed/MEDLINE from July 1st 2017 until September 30th 2021. Prognostic models published before April 1st 2017 were obtained from a previously conducted reviews.
The studies were considered eligible if they met the pre-established criteria as follows: (1) the article must describe either the development or external validation of a prediction model, or an update to a previously developed model; (2) the model in question must contain multiple predictors; (3) the predictors used must be routinely collected in Dutch Obstetric Care; (4) the predictors must be available and/or measured prior to 16 weeks and 0 days of gestation; (5) The study population must comprise pregnant women; (6) the model must be based on weighted risk predictors that have been identified through multivariate analysis; and (7) the model must be used to predict the GDM outcomes.
Data was collected through a pre-determined data extraction form, which included specific items related to study type, domain, outcome, development and validation, model performance (measured by AUROC), risk of bias, and applicability. To evaluate the risk of bias and applicability, the prediction model risk of bias assessment tool (PROBAST) was utilized. The systematic review adhered to the PRISMA guidelines for reporting.
RESULTS: In this study, 24 studies on GDM prediction models were selected for analysis, with 12 studies retrieved from the latest databases after rigorous selection. The final analysis included 20 models developed for GDM and 57 models externally validated for GDM. The developing models demonstrated AUROC values ranging from 0.64 to 0.88 (mean), but their performance in external validation studies was slightly lower, with AUROC values ranging from 0.60 to 0.87 (mean). Compared to all other models evaluated, Nanda's model has demonstrated a relatively stable performance in terms of AUROC values across both self-validation and external validation (0.73-0.79). Which suggests that Nanda's model may be more reliable and consistent in predicting outcomes compared to other models.
CONCLUSIONS: By analyzing and summarizing existing literature, this study presents information on current GDM prediction models with maternal characteristics and offers suggestions on the selection of predictors for future models, serving as a useful reference for future model development and enhancement. The study's findings suggest that although the model developed by Nanda et al. shows promise for predicting

GDM, the other models exhibit lower performance. As technology and research improve, we expect better GDM prediction models in the future.

Introduction

Gestational diabetes mellitus (GDM)[1] is frequently occurring pregnancy complication, with a globally standardized prevalence rate of 14.0%[2]. It increases the risk of complications during pregnancy and delivery, as well as raise the possibility of developing type 2 diabetes mellitus (DMII), cardiovascular disease and metabolic syndromes in the future[3,4]. Additionally, offspring of mothers with GDM are at increased risk of complications such as asthma, wheeze[5,6] obesity and abnormal glucose metabolism in mid-childhood[7].

The screening method for diagnosing GDM lacks international consensus and varies by healthcare provider and time[3]. Screening approaches include universal or selective screening based on certain risk factors or a combination of both[3]. Moreover, the diagnosis of GDM is complicated by the lack of a "gold standard" test. While the oral glucose tolerance test (OGTT) is commonly used, the optimal diagnostic criteria for the OGTT vary among different stakeholders[3,8]. Consequently, many women do not receive the appropriate diabetes screening. However, prediction models for GDM offer a novel approach to identifying high-risk women without extensive screening, which reduces the unnecessary burdens on pregnant women. By using these models, no cases are overlooked, ensuring that all at-risk individuals receive the appropriate attention and care.

Numerous studies have identified a range of factors contributing to the increased risk of gestational diabetes mellitus (GDM). Some research has established a significant association between higher maternal age and BMI with an increased risk of GDM[9,10]. Additional factors, including maternal history, ethnicity, prior GDM, and a family history of diabetes, have also been demonstrated to be significant GDM risk factors[11]. Bar-Zeev et al. found that for women who smoked equal or greater amounts of cigarettes, higher odds of GDM were observed across all BMI categories and gestational weight gain, compared to pregnant women who did not smoke[4]. Furthermore, a recent systematic review revealed a significant association between GDM and environmental factors, including air pollution, climate factors, chemicals and metals[7]. Studies by Tang et al.[12] and Miron-Celis et al.[13] have found that exposure to air pollution such as particulate matter with a diameter ≤ 2.5 μm ($PM_{2.5}$) and nitrogen dioxide (NO_2) are significantly associated with an increased risk of GDM based on existing evidence.
Several studies have shown that proactive identification of risk factors and timely preventive interventions such as applying physical activity, diet and probiotic intervention, might be able to lower the probability of GDM[1,14,15]. As such, early prediction of GDM is vital for effective risk management, and for improving maternal and neonatal outcomes. As a result, an increasing number of prediction models in the

field of obstetrics are being formulated [16].

Despite the significant increase in publications related to prognostic models for GDM, a consensus has yet to be reached on the best model for predicting the onset of GDM. From 2018 to 2020, Meertens et al.[17–20] published four systematic reviews (EXPECT studies) for first trimester obstetric prediction models for four different adverse pregnancy outcomes (GDM, PE, sPTB and SGA).Since then, no updates have been made, and there remains no agreement on the best prediction model for GDM.

To address this knowledge gap, the main study planned to conduct an up-to-date systematic review of existing models in obstetrics. This study is a sub-study of the larger research only focusing on updating the prediction model for GDM. In this systematic review, our objective is to present an updated review of existing first trimester prediction models for GDM based on maternal characteristics. By examining the current landscape of these models, we seek to uncover gaps or limitations in the research, which offer initial insights for further model research.

Methods

As previously mentioned in the introduction, the current study is a sub-analysis of the main study, focusing solely on studies related to the GDM model. Therefore, the analysis and summary were mainly aimed at information on the GDM prediction model. The methodology for this systematic review was documented and registered in the PROSPERO database, in accordance with established systematic review protocols.

Eligibility Criteria

A thorough search on the scientific databases PubMed and MEDLINE from July 1st, 2017 until September 30th, 2021 was performed to identify relevant studies. The studies published before April 1st, 2017 were sourced from the EXPECT study conducted by Meertens et al[17]. The predetermined in-and exclusion criteria are listed in Table 1.

Table 1 Criteria to Guide the Literature Search Following the PROSPERO protocol.

Inclusion Criteria	Exclusion Criteria
All forms of study designs are eligible for inclusion.	Reviews, letters, communications, editorials, and Case reports Not published in English.
The article must describe the development or external validation of a prediction model, or an update to a previously developed model.	The model is not a prediction model
The model in question should contain multiple predictors, which must have been routinely collected in Dutch Obstetric Care, and are available and/or measured before 16 weeks and 0 days of gestation	The models' predictors focus solely on biomarkers. Not first trimester model
The study population comprises pregnant women.	Wrong population that does not

	align with the research question, such as non-pregnant women or men, children, or individuals with irrelevant medical conditions.
The model must be based on weighted risk predictors that have been identified through multivariate analysis, and it is supposed used to predict GDM outcomes	Outcomes that did not align with our research question.

Search strategy

The search strategy employed in this study was based on the one used by Meertens et al.[17–20], we used a set of keywords related to prediction studies such as "predictive model", "prediction", "risk calculator", "risk model", "risk score", "algorithm", "risk assessment", "nomogram", "prognostic model", "scoring system", "screening model", "decision rule", "Pregnancy complications", "First trimester". Along with synonyms for outcomes such as "Gestational diabetes mellitus (GDM)", "Gestational diabetes", "pregnancy induced diabetes", "pregnancy-induced diabetes" in the title, abstract, or MeSH terms to identify relevant articles. The chosen articles were brought into Mendeley, where any duplicate articles were removed.

Study selection

Two independent reviewers evaluated all articles based on their title and abstract, using predefined inclusion and exclusion criteria. Eligible articles had their full text retrieved and assessed by the review team. In case of any disagreements between the reviewers, the full text was discussed and differences were resolved by consensus or through mediation with one other author.

Data collection

Data collection was done independently by three reviewers. Data extraction was conducted separately for all the included models, allowing for a comprehensive evaluation of each model's strengths and weaknesses.

Eligible studies were further categorized into three distinct groups based on their study design: development studies, external validation studies, as well as development and external validation studies combined. Relevant items were extracted from each selected article by using a pre-specified data extraction form, that contains items related to study type, domain, outcome, development and validation, bias, and applicability. In cases where an article detailed the creation or external validation of existing models, data extraction was performed separately for each model. The performance of each model in predicting gestational diabetes mellitus (GDM) was evaluated based on their reported area under the receiver operating characteristic curve (AUROC) values and corresponding 95% confidence intervals.

Risk of bias and applicability (quality) assessment

Risk of bias and applicability assessment was performed using the prediction model risk of bias assessment tool (PROBAST). PROBAST has 4 domains (participants, predictors, outcome, and analysis) and 20 signaling questions to assess the risk of bias. Answering "yes" or "probably yes" to all questions in a domain means low risk of bias, while "no" or "probably no" on one or more questions indicates potential for bias. In terms of applicability assessment, the review question was assessed across three domains: participants, predictors, and outcomes. Each domain was rated as low, high, or unclear, with a low concern for applicability indicating a good match between the review question and the study.

The present report followed the systematic review reporting guidelines outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

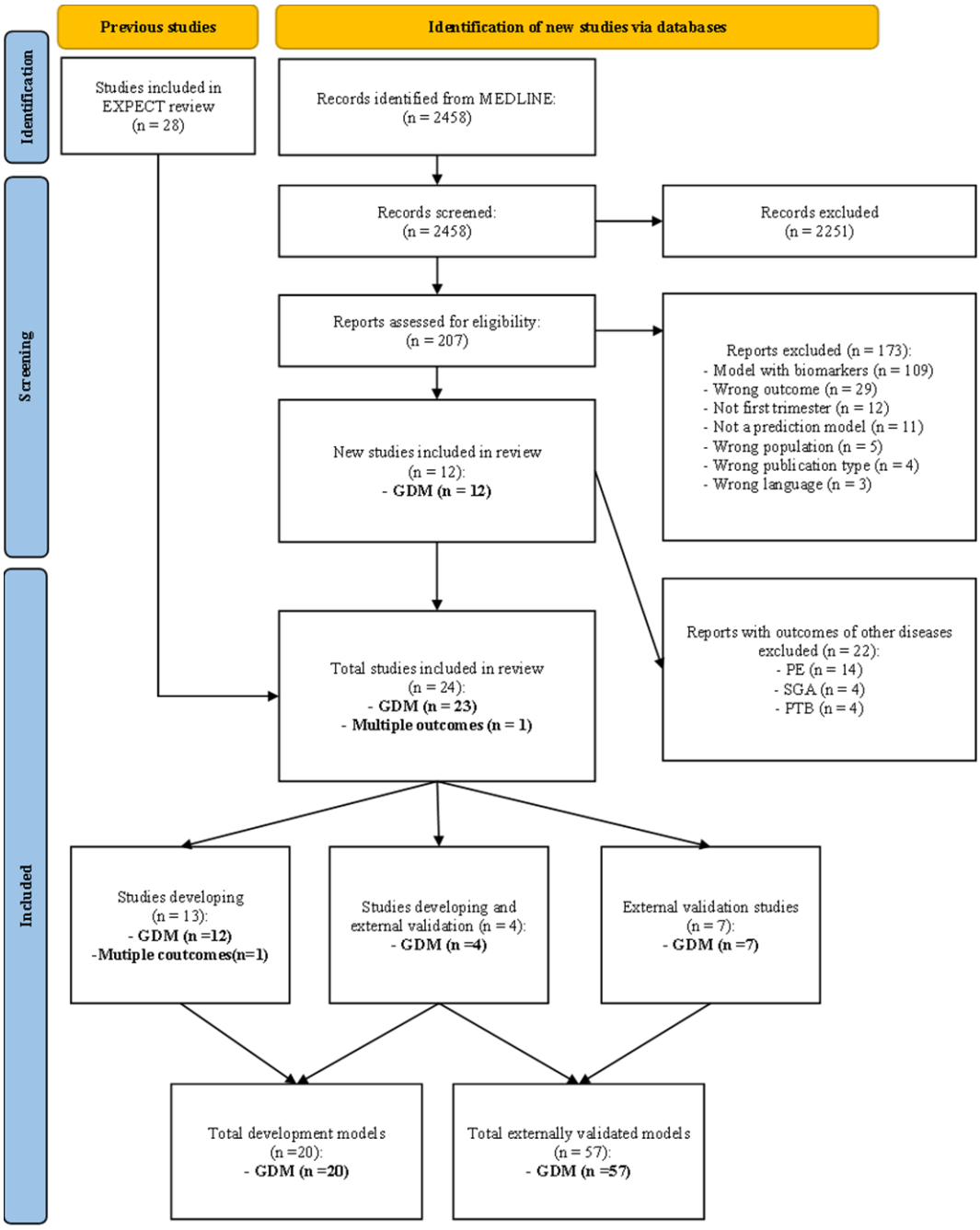
Results

Result of the Search

As depicted in Figure 1, the initial search yielded 2458 unique articles. Following title and abstract screening, 207 articles were identified as potentially relevant and were retrieved for further assessment. Of these 207 articles, 173 were subsequently excluded for the following reasons: focusing solely on biomarker models (n=109), having outcomes that did not align with our research question (n=29), not pertaining to first trimester models (n=12) or prediction models (n=11), having populations that did not meet our eligibility criteria (n=5), publication types that were not applicable to our criteria (n=4), and articles that were not written in English (n=3). Among the remaining 34 reports, 22 papers were excluded from this study as their models were not GDM prediction models. The remaining 12 reports were included in this research.

As highlighted in bold in Figure 1, a total of twenty-four studies on GDM prediction with maternal characteristics were identified, of which one study included models predicting multiple adverse outcomes such as GDM, PE, and SGA as outcomes. Twelves studies were retrieved from the latest databases after applying the selection criteria, and the other twelves studies were from a previous study by Meertens et al[17]. Among the 24 studies, 13 are development studies, 4 are development and external validation (combination) studies, and 7 are external validation studies. A total of 20 models developed for GDM and 57 models externally validated for GDM were included for analysis.

Figure 1 Flowchart of the study selection process.



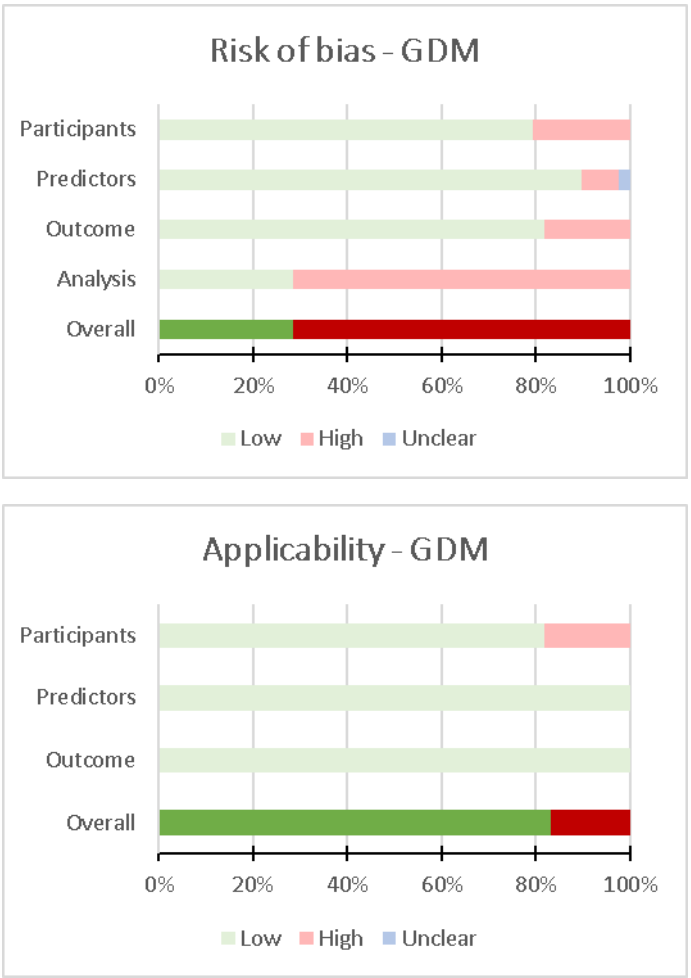
Risk of bias and applicability assessment

The results of risk of bias and applicability assessment are summarized in Figure 2. There is a low risk of bias in the participants, predictors, and outcomes domains. This suggests that the studies included in the review have a strong alignment with the review question and are relevant to the population of interest. Moreover, the predictors used in the studies are appropriate, and the study outcomes are meaningful in addressing the review question. On the other hand, most studies exhibited a high risk of bias in their analysis section, attributable to missing information on various aspects, such as the description of the performance measures of the model, missing data on predictors and outcomes, methods used for handling missing data, model development procedures, and more. To be more specific, numerous studies lacked information on how they

appropriately managed participants with missing data. Only two studies provided information on how they adequately considered data complexity. Most studies inappropriately evaluated model performance, and many did not account for model overfitting and optimism in model performance. Furthermore, a majority of the studies lacked information on whether the predictors and their assigned weights in the final model were consistent with the results of the reported multivariate analysis.

As shown in Figure 2, it is apparent that the concerns regarding applicability of most included prediction models for GDM are low. Low concerns regarding applicability, in this context, implies that the models are well-suited for the intended population and setting. It indicates that the study participants closely resemble the population of interest, the predictors align well with the review question, and the study outcomes are relevant and meaningful.

Figure 2 Risk of bias and applicability assessment of all studies by using PROBAST.



Characteristics of the included studies

Table 2 shows the key characteristics of the study designs, sample size, predictors, outcome, modeling method, and predictive performance of the included studies. The included studies on gestational diabetes mellitus (GDM) prediction models consisted of 13 development studies, 7 external validation studies, and 4 development and

external validation (combination) studies, and were published between 1997 and 2021. On average, each development study described 1.2 model, while external validation studies described approximately 5.9 models. Most studies used prospective cohorts and were conducted in various countries, with a concentration in Europe and North America. The number of participants used in model development varied from 134 to 771140, with a median of 1876. For external validation and combination studies, the number of participants ranged from 510 to 41577 (median 1266) and 980 to 75161 (median 5232.5), respectively.

The number of events in model development studies, ranged from 14 to 48608, with a median of 68. For external validation and combination studies, the number of events ranged from 47 to 381 (median 181) and 231 to 1827 (median 688.5), respectively.
The number of events per candidate predictor was calculated for all 17 development studies. Among them, 7 studies had a EPV less than 10, 2 studies had a EPV between 10 and 20, 7 studies had a EPV greater than 20, and for one study the EPV was unknown. The number of candidate predictors varied among studies, ranging from 7 to 26 (median 14) for development studies and 15 to 43 (median 20) for combination studies. The number of predictors in the final model ranged from 2 to 19 (median 7) for development studies, and 7 to 11, with a median of 7.5, for combination studies.

In all of the development studies, the logistic regression method was employed. Discrimination (C-statistic) was the primary method used to evaluate predictive performance, with a range of 0.60 to 0.88 for all studies. Some studies conducted internal validation, with 8 of the development studies and 3 of the combination studies reporting it. The method of validation varied among studies, with cross-validation, random split of data and nonrandom split of data being the most commonly used methods. However, it should be noted that 9 out of 13 development studies were not internally validated. Calibration was also reported for 7 out of 24 studies, with Hosmer-Lemeshow test and Calibration plot being the most common methods.

Table 2 Characteristics of studies related to the GDM model.

Item and Categories	GDM		
	Development studies	External validation studies	Combination studies
Number of studies	13	7	4
Year of publication (min-max)	1997-2020	2009-2021	2015-2020
No. models per study	1-4	1-9	Development:1 External validation: 2-6
Study type			
• Prospective	7 (54)	6 (86)	3 (75)
• Retrospective	4 (31)	0 (0)	0 (0)
• Other/unclear	2 (15)	1 (14)	1 (25)

Country			
• Africa	0 (0)	0 (0)	0 (0)
• Asia	3 (23)	0 (0)	1 (25)
• Europe	5 (38)	6 (86)	2 (50)
• North America	3 (23)	1 (14)	0 (0)
• Oceania	1 (8)	0 (0)	1 (25)
• South America	1 (8)	0 (0)	0 (0)
Sample size			
• No. participants	1876 (134-771140)	1266 (510-41577)	5232.5 (980-75161)
• No. events	68 (14-48608)	181(47-381)	688.5 (231-1827)
• Not reported	1	2	NA
Predictors			
• No. candidate predictors	14 (7-26)	0 (0-251)	20(15-43)
• No. predictors in final model	7 (2-19)	NA	7.5 (7-11)
EPV			
• EPV<10	6 (46)	0 (0)	1 (25)
• EPV 10-20	1 (8)	0 (0)	1 (25)
• EPV >20	6 (46)	0 (0)	1 (25)
• Not applicable	0 (0)	7 (100)	1 (25)
Modeling method			
• Logistic regression	13 (100)	0 (0)	4 (100)
• Not applicable	0 (0)	7 (100)	0 (0)
Predictive performance			
Discrimination			
• C-statistic, range	0.63-0.82	0.60-0.87	0.64-0.88
• Not reported	3 (23)	0 (0)	1 (25)
Calibration			
• Reported	2 (15)	4 (57)	1 (25)
• Hosmer-Lemeshow	2 (15)	0 (0)	0 (0)
• Calibration Plot	0 (0)	4 (57)	1 (25)
• Both discrimination and calibration reported	2 (15)	4 (57)	1 (25)
Internal validation			
Internally validated models	8 (62)	NA	3 (75)
Method of validation			
• Cross-validation	0 (0)	0 (0)	2 (50)
• Random split of data	2 (15)	0 (0)	0 (0)
• Nonrandom split of data	2 (15)	0 (0)	0 (0)
• Unclear	0 (0)	0 (0)	1 (25)

• Not applicable	9 (70)	7 (100)	1 (25)

Numbers are presented as n (%) or median, unless stated otherwise. NA means Not applicable

Receiver operating characteristic (ROC) of the included model development studies

As shown in Table 3, the predictors used in the models include previous GDM, family history of diabetes, maternal age, weight, BMI, history of smoking, among others. The AUROC values of the original models range from 0.64 to 0.88. However, their performance in external validation studies is slightly poorer, with AUROC values ranging from 0.60 to 0.87. Fourteen out of twenty models[21–33] have been validated in multiple studies, while others have only been validated once or not at all. Additionally, there is a high variation in the AUROC values among the external validation studies for the four model development studies conducted by Syngelaki et al.[22], Teede et al.[26], Van Leeuwen et al.[28], and Shirazian et al[29]. The model developed by Nanda demonstrates a more stable interval for the AUROC in comparison to other evaluated models. Furthermore, the model performs well in multiple external validations and holds the greatest potential for practical application.

Table 3 Predictive performance of models for gestational diabetes mellitus (GDM)

Model development studies	Predictors in model	AUROC (95% CI)		External validation studies
		Original publication	External Validations	
Sweeting et al, 2017[21]	Previous GDM, East Asian, South Asian, Family history of Diabetes, Parity, Maternal age, BMI	0.88 (0.85-0.92)	0.72 (0.67-0.77) 0.71(0.68-0.75)	Meertens et al, 2019[17] Kotzaeridi et al, 2021[34]
Donovan et al, 2019[35]	Ethnicity, maternal age at delivery, pre-pregnancy BMI, family history of diabetes, pre-existing hypertension	0.73 (0.73 - 0.74)	NA	NA
Benhalima et al, 2020[36]	A first degree relative with diabetes, a history of smoking before pregnancy, Asian	0.72 (0.69-0.76)	0.72 (0.68-0.75)	Kotzaeridi et al, 2021[34]

Model development studies	Predictors in model	AUROC (95% CI)		External validation studies
		Original publication	External Validations	
	origin, age, height, BMI, history of GDM			
Schaefer et al, 2018[37]	Maternal age, BMI, family history of diabetes, history of GDM, weight gain during pregnancy	0.64 (0.62-0.66)	NA	NA
Garmendia et al, 2020[38]	Maternal age, education level, marital status, parity, family history of type 2 diabetes, previous maternal hypertension, previous abnormal pregnancies, PE in previous pregnancy, GDM in previous pregnancy, smoking, alcohol consumption, pre pregnancy nutritional status	0.74 (no CI)	NA	NA
Syngelaki et al, 2015[22]	Previous history of GDM, family history of first/second degree relative with DM, maternal age, weight, height, racial origin, method of conception, birth	NA	0.87 (0.84–0.90) 0.68 (0.62-0.74) 0.72 (0.68-0.75)	Sweeting et al, 2017[21] Meertens et al, 2019[17] Kotzaeridi et al, 2021[34]

Model development studies	Predictors in model	AUROC (95% CI)		External validation studies
		Original publication	External Validations	
	weight z score of neonates in last pregnancy			
Eleftheriades et al, 2014[23]	Maternal weight, maternal age	0.73 (0.65 - 0.81)	NA	Schaefer et al, 2018[37]
			0.68 (0.63-0.73)	Meertens et al, 2019[17]
			Recalibrated models: 0.70 (0.65-0.74)	Lamain- de Ruiter et al, 2016[1]
Gabbay et al, 2014[24]	Maternal age, prior GDM, SBP, first trimester BMI, Race	0.82 (0.77 - 087)	0.72 (0.67-0.77)	Meertens et al, 2019[17]
			0.72 (0.68-0.75)	Kotzaeridi et al, 2021[34]
			Recalibrated models: 0.75 (0.71-0.79)	Lamain- de Ruiter et al, 2016[1]
Tran et al, 2013[25]	Maternal age, BMI	ADA: 0.71 (0.68-0.75)	0.70 (0.64-0.75)	Meertens et al, 2019[17]
		ADIPS 0.64 (0.62 - 0.67)	NA	Schaefer et al, 2018[37]
		IADPSG: 0.65 (0.62 - 0.67)	Recalibrated models: 0.67 (0.63- 0.72)	Lamain- de Ruiter et al, 2016[1]
		WHO: 0.63 (0.60 - 0.65)		
Teede et al, 2011[26]	Maternal age, BMI, ethnicity, fam history of DM, prior GDM, history of poor obstetric outcome	0.70(no CI)	0.73 (0.68-0.78)	Meertens et al, 2019[17]
			0.66(0.62-0.70)	Benhalima et al,2020
			0.69 (0.65-0.73)	Kotzaeridi et al, 2021[34]
			NA	Schaefer et al, 2018[37]
			Recalibrated models: 0.77(0.73-0.81)	Lamain- de Ruiter et al, 2016[1]
			0.77 (0.76-0.77)	Syngelaki et al, 2015[22]
			0.77	Huvinen et al, 2018[39]

Model development studies	Predictors in model	AUROC (95% CI)		External validation studies
		Original publication	External Validations	
			0.74 (0.70-0.78)	Thériault et al, 2014[40]
			0.60 (0.56-0.64)	Lovati et al, 2013[6]
Nanda et al, 2011[27]	Maternal age, BMI, ethnicity, parity, prior GDM, prior LGA	0.79 (0.76 - 0.82)	0.78 (0.75-0.82)	Sweeting et al, 2017
			0.75 (0.70-0.80)	Meertens et al, 2019[17]
			0.73 (0.69-0.77)	Kotzaeridi et al, 2021
			Recalibrated models: 0.78 (0.74-0.82)	Lamain- de Ruiter et al, 2016[1]
			0.79 (0.78-0.79)	Syngelaki et al, 2015[22]
Van Leeuwen et al, 2010[28]	Eethnicity, family history of GDM, multipara without history of GDM, multipara without history of GDM, BMI	0.77 (0.69 - 0.85)	0.64 (0.61-0.69)	Sweeting et al, 2017[21]
			0.74 (0.70-0.79)	Meertens et al, 2019[17]
			0.67 (0.63-0.71)	Benhalima et al, 2020
			0.71 (0.67-0.75)	Kotzaeridi et al, 2021[34]
			NA	Schaefer et al, 2018[37]
			Recalibrated models: 0.74 (0.7-0.78)	Lamain- de Ruiter et al, 2016[1]
			0.77 (0.77-0.78)	Syngelaki et al, 2015[22]
			0.74	Huvinen et al, 2018[39]
			0.76 (0.73-0.78)	Thériault et al, 2014[40]
Shirazian et al, 2009[29]	Maternal age, BMI, family history of DM type 2	NA	0.71 (0.66-0.76)	Meertens et al, 2019[17]
			0.61(0.57-0.65)	Kotzaeridi et al, 2021[34]
			NA	Schaefer et al, 2018[37]
			Recalibrated models:	Lamain- de

Model development studies	Predictors in model	AUROC (95% CI)		External validation studies
		Original publication	External Validations	
			0.71 (0.67-0.75)	Ruiter et al, 2016[1]
Phaloprakarn et al, 2009[30]	Maternal age, BMI, family history of diabetes, history of macrosomia, history of 2 or more abortions	0.77(0.75 - 0.79)	0.74 (0.69-0.79)	Meertens et al, 2019[17]
			0.68(0.64-0.71)	Kotzaeridi et al, 2021[34]
Naylor et al, 1997[30]	Maternal age, BMI, Race	0.68 (no CI)	0.68 (0.63-0.73)	Meertens et al, 2019[17]
			0.66(0.62-0.69)	Kotzaeridi et al, 2021[34]
			NA	Schaefer et al, 2018[37]
			Recalibrated models: 0.72 (0.68-0.76)	Lamain- de Ruiter et al, 2016[1]
			0.69 (0.68-0.69)	Syngelaki et al, 2015[22]
			0.67 (0.64-0.70)	Thériault et al, 2014[40]
			0.64 (0.56-0.72)	van Leeuwen et al, 2009[41]
Caliskan et al, 2004[32]	Maternal age > or equal to 25, BMI > or equal to 25, diabetes in first degree relative, prior macrosomic infant, history of adverse obstetric outcome (either recurrent spontaneous abortions, fetal anomaly with normal karyotype, prior unexplained in	NA	0.65(0.61-0.68)	Kotzaeridi et al, 2021[34]
			Recalibrated models: 0.73 (0.69-0.76)	Lamain- de Ruiter et al, 2016[1]
			0.70 (0.70-0.70)	Syngelaki et al, 2015[22]
			0.68 (0.65-0.71)	Thériault et al, 2014[40]

Model development studies	Predictors in model	AUROC (95% CI)		External validation studies
		Original publication	External Validations	
	utero fetal death at GA > or equal to 20 weeks of GA)			
Syngelaki et al, 2011[33]	Previous history of GDM, family history of first/second degree relative with DM, maternal age, weight, height, racial origin, method of conception, birth weight z score of neonates in last pregnancy	NA	0.75 (0.70-0.80) Recalibrated models: 0.71 (0.66-0.75)	Meertens et al, 2019[17] Lamain- de Ruiter et al, 2016[1]

Discussion

In our review of 24 studies, we analyzed the performance of 20 prediction models for GDM risk. Most models had moderate discriminative performance, with an AUROC of approximately 0.7 (mean). Based on the analysis of the models, the AUROC values of model conducted by Nanda et al.[27] remaining above 0.75 during self-validation and external validation in most cases. Moreover, this model demonstrated less variability in the range of AUROC values compared to other models with high AUROC values. These results indicate that Nanda et al.[27] model showed the most promise for predictive performance, which aligns with the findings of Meertens et al.[17], but differs from the results reported by Huvinen et al[39].

Implications of the results for practice, policy, and future research

There is currently no agreement on the most suitable model for clinical practice. However, among the 20 models included in this study, the model developed by Nanda et al.[27] demonstrated relatively good discrimination compared to other models. While the development and validation of GDM prediction models are still in their early stages, these findings provide valuable insights into the potential clinical applications of GDM prediction models. Moreover, these findings highlight the need for further research and validation in real-world settings to improve the accuracy and generalization performance of GDM prediction models.

Although the model developed by Nanda et al.[27] showed relatively good discrimination, it is worth noting that there is currently no established gold standard for the AUROC threshold required for clinical application. Some studies suggest that an AUROC above 0.8 is considered excellent performance, while others argue that an AUROC of 0.9 or higher is required to be considered perfect performance[42,43]. Even though additional performance measures are used to evaluate models alongside AUROC, we feel that a consensus on the benchmark for model assessment could enhance our ability to pinpoint the most optimal models.

It is worth noting that all the studies included in this review utilized logic regression methods in developing their GDM prediction models. Recently, Chan et al.[44] have introduced a novel machine learning technique to predict GDM early on using the elemental composition in fingernails, presenting a high AUROC, with a value of 0.81. This promising result suggests that innovative approaches such as machine learning and deep learning methods may hold potential for improving the accuracy and generalization performance of GDM prediction models. However, it is important to consider that machine learning approaches may not always offer a significant advantage over traditional statistical methods. In a previous study by Kuhle et al.[45], it was found that machine learning approaches did not offer a substantial advantage over logistic regression when predicting fetal growth abnormalities.

Another important point to highlight is that there is mounting evidence indicating that, in addition to traditional risk factors, environmental factors such as $PM_{2.5}$, smoking etc., may play a role in the development of GDM[13,46]. A small number of included studies in this review have integrated home environmental factors such as smoking into model predictors[36,38], but both studies lack external validation. Despite conducting some additional searches for studies on prediction models that incorporate environmental factors, we found a dearth of research in this area. This lack of incorporation may be due to a shortage of relevant data and quantitative indicators as well as a tendency to overlook environmental factors in clinical research. However, from a One Health perspective, the combination of human health and environmental health in research is considered essential to achieve the best health for everyone[47]. The One Health approach recognizes that human health, animal health, and environmental health are interdependent and connected. Therefore, we suggest that future research on GDM prediction models could consider incorporating environmental factors, such as $PM_{2.5}$ and traffic air pollutant NO_2, which have been identified as risk factors for GDM in multiple studies[7,12,13].

Strengths and Limitations

One of the strengths of our study is that it is an up-to-date and comprehensive evaluation of the available evidence on GDM prediction models based on maternal characteristics, incorporating the latest research and developments in the field. This study represents the most recent version of a systematic review of existing prediction models for the risk of GDM, enhancing the accuracy and reliability of our findings.

In addition to summarizing the accuracy and generalizability of different GDM prediction models, we also included a quality assessment of the included studies to

evaluate the risk of bias and applicability of the studies in our review. By including a quality assessment, we were able to provide a more comprehensive evaluation of the available evidence.

As an update to the review conducted by Meertens et al.[17], we have excluded prediction models that used biochemical variables as risk predictors due to practical considerations such as feasibility and cost-effectiveness in clinical settings. The systematic review only includes models that use maternal characteristics that were routinely collected in Dutch Obstetric Care as predictors. However, it is important to note that the scope of this study may be limited as more studies emerge on models that incorporate biomarkers. In fact, models like the one developed by Benhalima et al.[36], which utilize biomarkers such as fasting plasma glucose (FPG), has demonstrated superior performance compared to models that exclude such biomarkers as predictors[34].

In addition to the limitation discussed above, there are several other limitations to this study. A high risk of bias exists in many studies due to various factors, including insufficient information about model development procedures, inadequate handling of missing data and other issues. One possible explanation for this is the time gap between the development of GDM prediction models and the introduction of the PROBAST guidelines, as it may lead to a lack of crucial information for conducting a comprehensive bias assessment. For example, information on whether the predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis was often not reported, leading to high risk of bias in the analysis domain across most of models. Another limitation is the lack of studies from the continent of Africa. While studies from other regions have provided valuable insights into GDM risk factors and prediction models, the epidemiology of GDM in Africa may differ significantly from other regions, due to differences in genetic, economic situations and environmental factors. Therefore, it is important to give more attention to the development and validation of prediction model studies in Africa in future research. Finally, there are four types of diagnostic criteria for GDM: American Diabetes Association (ADA), Australasian Diabetes in Pregnancy Society (ADIPS), International Association of the Diabetes and Pregnancy Study Groups (IADPSG), and World Health Organization (WHO). Different diagnostic criteria were used in the predictive models included in a systematic review. Which may reduce the comparability of the included studies and limit the credibility of the conclusions.

Conclusion

Through a comprehensive review of existing literature, this study provides a summary of the current state of GDM prediction models that based on maternal characteristics, nd offers recommendations on potential new predictors, such as NO_2. Furthermore, the identified limitations in this study can inform the development of future prediction models, serving as a valuable resource for advancing future model development and refinement. Researchers aiming to develop their own prediction models can use these insights as a reference for model development and predictor selection.

Among the currently included GDM prediction models, the one conducted by Nanda et al.[27] has demonstrated promising potential for clinical application. However, as model

building techniques continue to evolve and improve, such as using machine learning and AI, along with the publication of more research on disease-related predictors, it is expected that GDM prediction models will become more accurate and generalizable.

Reference

1. Ruiter ML de, Kwee A, Naaktgeboren CA, et al. External validation of prognostic models to predict risk of gestational diabetes mellitus in one Dutch cohort: prospective multicentre cohort study. *BMJ*. 2016;354:i4338. doi:10.1136/bmj.i4338

2. Wang H, Li N, Chivese T, et al. IDF Diabetes Atlas: Estimation of Global and Regional Gestational Diabetes Mellitus Prevalence for 2021 by International Association of Diabetes in Pregnancy Study Group's Criteria. *Diabetes Res Clin Pract*. 2022;183:109050. doi:10.1016/j.diabres.2021.109050

3. Zhu Y, Zhang C. Prevalence of Gestational Diabetes and Risk of Progression to Type 2 Diabetes: a Global Perspective. *Curr Diab Rep*. 2016;16(1):7. doi:10.1007/s11892-015-0699-x

4. Bar-Zeev Y, Haile ZT, Chertok IA. Association Between Prenatal Smoking and Gestational Diabetes Mellitus. *Obstet Gynecol*. 2020;135(1):91. doi:10.1097/AOG.0000000000003602

5. Adgent MA, Gebretsadik T, Reedus J, et al. Gestational diabetes and childhood asthma in a racially diverse U.S. pregnancy cohort. *Pediatr Allergy Immunol Off Publ Eur Soc Pediatr Allergy Immunol*. 2021;32(6):1190-1196. doi:10.1111/pai.13523

6. Lovati E, Beneventi F, Simonetta M, et al. Gestational diabetes mellitus: Including serum pregnancy-associated plasma protein-A testing in the clinical management of primiparous women? A case–control study. *Diabetes Res Clin Pract*. 2013;100(3):340-347. doi:10.1016/j.diabres.2013.04.002

7. Eberle C, Stichling S. Environmental health influences in pregnancy and risk of gestational diabetes mellitus: a systematic review. *BMC Public Health*. 2022;22(1):1572. doi:10.1186/s12889-022-13965-5

8. Hanson E, Ringmets I, Kirss A, Laan M, Rull K. Screening of Gestational Diabetes and Its Risk Factors: Pregnancy Outcome of Women with Gestational Diabetes Risk Factors According to Glycose Tolerance Test Results. *J Clin Med*. 2022;11(17):4953. doi:10.3390/jcm11174953

9. Yong HY, Mohd Shariff Z, Mohd Yusof BN, et al. Independent and combined effects of age, body mass index and gestational weight gain on the risk of gestational diabetes mellitus. *Sci Rep*. 2020;10(1):8486. doi:10.1038/s41598-020-65251-2

10. Zheng QX, Wang HW, Jiang XM, et al. Prepregnancy body mass index and gestational weight gain are associated with maternal and infant adverse outcomes in Chinese women with gestational diabetes. *Sci Rep*. 2022;12(1):2749. doi:10.1038/s41598-022-06733-3

11. McIntyre HD, Catalano P, Zhang C, Desoye G, Mathiesen ER, Damm P. Gestational diabetes mellitus. *Nat Rev Dis Primer*. 2019;5(1):47. doi:10.1038/s41572-019-0098-8

12. Tang X, Zhou JB, Luo F, et al. Air pollution and gestational diabetes mellitus: evidence from cohort studies. *BMJ Open Diabetes Res Care*. 2020;8(1):e000937. doi:10.1136/bmjdrc-2019-000937

13. Miron-Celis M, Talarico R, Villeneuve PJ, et al. Critical windows of exposure to air pollution and gestational diabetes: assessing effect modification by maternal pre-existing conditions and environmental factors. *Environ Health*. 2023;22(1):26. doi:10.1186/s12940-023-00974-z

14. Koivusalo SB, Rönö K, Klemetti MM, et al. Gestational Diabetes Mellitus Can Be Prevented by Lifestyle Intervention: The Finnish Gestational Diabetes Prevention Study (RADIEL): A Randomized Controlled Trial. *Diabetes Care*. 2016;39(1):24-30. doi:10.2337/dc15-0511

15. Zakaria H, Abusanana S, Mussa BM, et al. The Role of Lifestyle Interventions in the Prevention and Treatment of Gestational Diabetes Mellitus. *Medicina (Mex)*. 2023;59(2):287. doi:10.3390/medicina59020287

16. Kleinrouweler CE, Cheong-See FM, Collins GS, et al. Prognostic models in obstetrics: available, but far from applicable. *Am J Obstet Gynecol*. 2016;214(1):79-90.e36. doi:10.1016/j.ajog.2015.06.013

17. Meertens LJE, Scheepers HCJ, van Kuijk SMJ, et al. External validation and clinical utility of prognostic prediction models for gestational diabetes mellitus: A prospective cohort study. *Acta Obstet Gynecol Scand*. 2020;99(7):891-900. doi:10.1111/aogs.13811

18. Meertens LJE, Scheepers HCJ, van Kuijk SMJ, et al. External Validation and Clinical Usefulness of First Trimester Prediction Models for the Risk of Preeclampsia: A Prospective Cohort Study. *Fetal Diagn Ther*. 2019;45(6):381-393. doi:10.1159/000490385

19. Meertens LJE, van Montfort P, Scheepers HCJ, et al. Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. *Acta Obstet Gynecol Scand*. 2018;97(8):907-920. doi:10.1111/aogs.13358

20. Meertens L, Smits L, van Kuijk S, et al. External validation and clinical usefulness

of first-trimester prediction models for small- and large-for-gestational-age infants: a prospective cohort study. *BJOG Int J Obstet Gynaecol*. 2019;126(4):472-484. doi:10.1111/1471-0528.15516

21. Sweeting AN, Appelblom H, Ross GP, et al. First trimester prediction of gestational diabetes mellitus: A clinical model based on maternal demographic parameters. *Diabetes Res Clin Pract*. 2017;127:44-50. doi:10.1016/j.diabres.2017.02.036

22. Syngelaki A, Pastides A, Kotecha R, Wright A, Akolekar R, Nicolaides KH. First-Trimester Screening for Gestational Diabetes Mellitus Based on Maternal Characteristics and History. *Fetal Diagn Ther*. 2015;38(1):14-21. doi:10.1159/000369970

23. Eleftheriades M, Papastefanou I, Lambrinoudaki I, et al. Elevated placental growth factor concentrations at 11–14weeks of gestation to predict gestational diabetes mellitus. *Metabolism*. 2014;63(11):1419-1425. doi:10.1016/j.metabol.2014.07.016

24. Gabbay-Benziv R, Doyle LE, Blitzer M, Baschat AA. First trimester prediction of maternal glycemic status. *J Perinat Med*. 2015;43(3):283-289. doi:10.1515/jpm-2014-0149

25. Tran TS, Hirst JE, Do MAT, Morris JM, Jeffery HE. Early Prediction of Gestational Diabetes Mellitus in Vietnam. *Diabetes Care*. 2013;36(3):618-624. doi:10.2337/dc12-1418

26. Teede HJ, Harrison CL, Teh WT, Paul E, Allan CA. Gestational diabetes: Development of an early risk prediction tool to facilitate opportunities for prevention. *Aust N Z J Obstet Gynaecol*. 2011;51(6):499-504. doi:10.1111/j.1479-828X.2011.01356.x

27. Nanda S, Savvidou M, Syngelaki A, Akolekar R, Nicolaides KH. Prediction of gestational diabetes mellitus by maternal factors and biomarkers at 11 to 13 weeks. *Prenat Diagn*. 2011;31(2):135-141. doi:10.1002/pd.2636

28. van Leeuwen M, Opmeer BC, Zweers EJK, et al. Estimating the risk of gestational diabetes mellitus: a clinical prediction model based on patient characteristics and medical history. *BJOG Int J Obstet Gynaecol*. 2010;117(1):69-75. doi:10.1111/j.1471-0528.2009.02425.x

29. Shirazian N, Emdadi R, Mahboubi M, et al. Screening for gestational diabetes: usefulness of clinical risk factors. *Arch Gynecol Obstet*. 2009;280(6):933-937. doi:10.1007/s00404-009-1027-y

30. Phaloprakarn C, Tangjitgamol S, Manusirivithaya S. A risk score for selective screening for gestational diabetes mellitus. *Eur J Obstet Gynecol Reprod Biol*. 2009;145(1):71-75. doi:10.1016/j.ejogrb.2009.04.016

31. Naylor CD, Sermer M, Chen E, Farine D. Selective Screening for Gestational Diabetes Mellitus. *N Engl J Med*. 1997;337(22):1591-1596. doi:10.1056/NEJM199711273372204

32. Caliskan E, Kayikcioglu F, Öztürk N, Koc S, Haberal A. A population-based risk factor scoring will decrease unnecessary testing for the diagnosis of gestational diabetes mellitus. *Acta Obstet Gynecol Scand*. 2004;83(6):524-530. doi:10.1111/j.0001-6349.2004.00389.x

33. Syngelaki A, Bredaki FE, Vaikousi E, Maiz N, Nicolaides KH. Body mass index at 11-13 weeks' gestation and pregnancy complications. *Fetal Diagn Ther*. 2011;30(4):250-265. doi:10.1159/000328083

34. Kotzaeridi G, Blätter J, Eppel D, et al. Performance of early risk assessment tools to predict the later development of gestational diabetes. *Eur J Clin Invest*. 2021;51(12):e13630. doi:10.1111/eci.13630

35. Donovan BM, Breheny PJ, Robinson JG, et al. Development and validation of a clinical model for preconception and early pregnancy risk prediction of gestational diabetes mellitus in nulliparous women. *PLOS ONE*. 2019;14(4):e0215173. doi:10.1371/journal.pone.0215173

36. Benhalima K, Van Crombrugge P, Moyson C, et al. Estimating the risk of gestational diabetes mellitus based on the 2013 WHO criteria: a prediction model based on clinical and biochemical variables in early pregnancy. *Acta Diabetol*. 2020;57(6):661-671. doi:10.1007/s00592-019-01469-5

37. Schaefer KK, Xiao W, Chen Q, et al. Prediction of gestational diabetes mellitus in the Born in Guangzhou Cohort Study, China. *Int J Gynecol Obstet*. 2018;143(2):164-171. doi:10.1002/ijgo.12627

38. Garmendia ML, Mondschein S, Montiel B, Kusanovic JP. Trends and predictors of gestational diabetes mellitus in Chile. *Int J Gynecol Obstet*. 2020;148(2):210-218. doi:10.1002/ijgo.13023

39. Huvinen E, Eriksson JG, Stach-Lempinen B, Tiitinen A, Koivusalo SB. Heterogeneity of gestational diabetes (GDM) and challenges in developing a GDM risk score. *Acta Diabetol*. 2018;55(12):1251-1259. doi:10.1007/s00592-018-1224-x

40. Thériault S, Forest JC, Massé J, Giguère Y. Validation of early risk-prediction models for gestational diabetes based on clinical characteristics. *Diabetes Res Clin Pract*. 2014;103(3):419-425. doi:10.1016/j.diabres.2013.12.009

41. van Leeuwen M, Opmeer BC, Zweers EJK, et al. External validation of a clinical scoring system for the risk of gestational diabetes mellitus. *Diabetes Res Clin Pract*. 2009;85(1):96-101. doi:10.1016/j.diabres.2009.04.025

42. Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *J Thorac Oncol*. 2010;5(9):1315-1316. doi:10.1097/JTO.0b013e3181ec173d

43. El Khouli RH, Macura KJ, Barker PB, et al. The Relationship of Temporal Resolution to Diagnostic Performance for Dynamic Contrast Enhanced (DCE) MRI of the Breast. *J Magn Reson Imaging JMRI*. 2009;30(5):999-1004. doi:10.1002/jmri.21947

44. Chan YN, Wang P, Chun KH, et al. A machine learning approach for early prediction of gestational diabetes mellitus using elemental contents in fingernails. *Sci Rep*. 2023;13(1):4184. doi:10.1038/s41598-023-31270-y

45. Kuhle S, Maguire B, Zhang H, et al. Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. *BMC Pregnancy Childbirth*. 2018;18(1):333. doi:10.1186/s12884-018-1971-2

46. Carroll X, Liang X, Zhang W, et al. Socioeconomic, environmental and lifestyle factors associated with gestational diabetes mellitus: A matched case-control study in Beijing, China. *Sci Rep*. 2018;8(1):8103. doi:10.1038/s41598-018-26412-6

47. Europe WHORO for. A health perspective on the role of the environment in One Health. In: *A Health Perspective on the Role of the Environment in One Health*. ; 2022. Accessed April 28, 2023. https://apps.who.int/iris/handle/10665/354574