ORIGINAL ARTICLE

WILEY

Performance of early risk assessment tools to predict the later development of gestational diabetes

Grammata Kotzaeridi[1] | Julia Blätter[1] | Daniel Eppel[1] | Ingo Rosicky[1] |
Martina Mittlböck[2] | Gülen Yerlikaya-Schatten[1] | Christian Schatten[1] | Peter Husslein[1] |
Wolfgang Eppel[1] | Evelyn A. Huhn[3] | Andrea Tura[4] | Christian S. Göbl[1] 🆔

[1]Department of Obstetrics and
Gynaecology, Medical University of
Vienna, Vienna, Austria

[2]Center of Medical Statistics, Informatics,
and Intelligent Systems, Section for Clinical
Biometrics, Medical University of Vienna,
Vienna, Austria

[3]Department of Obstetrics and
Gynaecology, University Hospital Basel,
Basel, Switzerland

[4]Metabolic Unit, CNR Institute of
Neuroscience, Padova, Italy

Correspondence
Christian S. Göbl, Department of Obstetrics
and Gynaecology, Medical University of
Vienna, Waehringer Guertel 18-20, A-1090
Vienna, Austria.
Email: christian.goebl@meduniwien.ac.at

Funding information
This study did not receive specific funding

Abstract

Background: Several prognostic models for gestational diabetes mellitus (GDM) are provided in the literature; however, their clinical significance has not been thoroughly evaluated, especially with regard to application at early gestation and in accordance with the most recent diagnostic criteria. This external validation study aimed to assess the predictive accuracy of published risk estimation models for the later development of GDM at early pregnancy.

Methods: In this cohort study, we prospectively included 1132 pregnant women. Risk evaluation was performed before 16 + 0 weeks of gestation including a routine laboratory examination. Study participants were followed-up until delivery to assess GDM status according to the IADPSG 2010 diagnostic criteria. Fifteen clinical prediction models were calculated according to the published literature.

Results: Gestational diabetes mellitus was diagnosed in 239 women, that is 21.1% of the study participants. Discrimination was assessed by the area under the ROC curve and ranged between 60.7% and 76.9%, corresponding to an acceptable accuracy. With some exceptions, calibration performance was poor as most models were developed based on older diagnostic criteria with lower prevalence and therefore tended to underestimate the risk of GDM. The highest variable importance scores were observed for history of GDM and routine laboratory parameters.

Conclusions: Most prediction models showed acceptable accuracy in terms of discrimination but lacked in calibration, which was strongly dependent on study settings. Simple biochemical variables such as fasting glucose, HbA1c and triglycerides can improve risk prediction. One model consisting of clinical and laboratory parameters showed satisfactory accuracy and could be used for further investigations.

KEYWORDS

early gestation, gestational diabetes mellitus, risk prediction model

1 | INTRODUCTION

According to the most recent WHO recommendations, gestational diabetes mellitus (GDM) is diagnosed in the late second or early third trimester by using a 75 g oral glucose tolerance test (OGTT).[1] Although there is no consensus about screening algorithms and diagnostic criteria for GDM before 24 weeks of gestation, risk stratification at early pregnancy could be beneficial to reduce diabetes associated co-morbidities due to timely interventions such as physical exercise and dietary changes.[2] For this purpose, anamnestic risk factors (maternal age, overweight or obesity, ethnicity, history of GDM in previous pregnancy and others) can be used to distinguish women with low and high risk for the later development of hyperglycaemia.[3] Although risk factor-based screening is criticized to have limited diagnostic accuracy, it has the advantage of cost-effectiveness and simplicity as clinical risk factors are easily obtainable from the patient's history already at early gestation.[4] Recently, some authors suggested that risk factor-based screening could be considerably improved by use of clinical prediction models consisting of statistical combinations of several GDM associated risk factors.[5] Their predictive performance could possibly be further improved by inclusion of fasting plasma glucose and other laboratory measurements such as triglycerides or glycated haemoglobin A1c (HbA1c).[6] As a consequence, various risk estimation tools have been published in the previous years. However, most of these previous studies used different GDM diagnosis criteria; therefore, their clinical benefit when applied to the most recent GDM definition is not well investigated yet.

This study aims to assess the performance of published low-invasive prediction models (ie risk estimation containing anamnestic data and routinely available fasting laboratory measurements) to predict the later development of GDM as defined by the recent WHO recommendations[1] as primary objective. Discrimination and calibration of several models, as well as the importance of included variables, were assessed in an independent population, and their possible clinical benefit was investigated and discussed.

2 | METHODS

2.1 | Study design and participants

In this prospective cohort study, we included 1132 pregnant women. A flow chart with detailed information about included and excluded patients is provided in the supplemental material (Figure S1). Study participants (≥18 years) were consecutively recruited among pregnant women attending our outpatient clinic (Department of Obstetrics and Gynaecology, Division of Obstetrics and Feto-Maternal Medicine, Medical University of Vienna) between January

2016 and July 2019. Women with pre-existing diabetes as well as women with history of bariatric surgery were excluded. A broad risk evaluation was performed before 16 + 0 weeks of gestation, including the assessment of pre-gestational body mass index (BMI, based on the self-reported pre-gestational weight), BMI at initial contact, maternal age, parity, obstetric history, family history of diabetes (first and/or second degree) and history of GDM in previous pregnancies. A detailed assessment of ethnicity was required for some prediction models and included the following ethnic categories: Caucasian (n = 881, 77.8%), African/Black (n = 47, 4.2%), Middle Eastern and North African (n = 76, 6.7%), Asian (n = 118, 10.4%), Latin American (n = 9, 0.8%) and Caribbean (n = 1, 0.1%). Due to reported differences in GDM risk between different Asian ethnic groups, the ethnic category 'Asian' was further subdivided into the subcategories South Asian (n = 82, 7.2%), Southeast Asian (n = 18, 1.6%), East Asian (n = 11, 1.0%) and Central Asian (n = 7, 0.6%). Obstetric history included adverse obstetric outcomes such as unexplained foetal death at >20 weeks of gestation, spontaneous abortions and foetal anomalies despite normal karyotype. To match the definitions used in certain prediction models, macrosomia in previous gestation was defined as birthweight greater than 4000 g, whereas large for gestational age (LGA) was defined as birthweight greater than 90th percentile using INTERGROWTH-21st newborn growth standards, which were also used to determine the birth weight z-scores.[7] In addition, a routine laboratory examination was performed at baseline visit with assessments of plasma glucose, triglycerides, glycated haemoglobin A1c (HbA1c) and urine glucose after at least 8 hours of fasting. Participants with fasting plasma glucose or HbA1c equal or exceeding 126 mg/dL or 6.5%, respectively, were excluded from the cohort. Thereafter, participants were followed-up until delivery to assess glycaemic status by using a 75 g 2 h OGTT according to the International Diabetes in Pregnancy Study Groups (IADPSG) recommendations[8] in a universal screening. In cases where fasting glucose was ≥92 mg/dL before 24 weeks of gestation, presence of GDM was verified by either early OGTT (n = 33) or self-monitored blood glucose (n = 25) in accordance with our national guidelines.[9] The majority of those patients had a history of GDM in previous pregnancies and showed elevated blood glucose levels throughout their entire current pregnancy, often requiring pharmacological intervention with insulin and/or metformin. Women with negative OGTT results, who received glucose lowering medication during follow-up due to macrosomia and after confirmation of hyperglycaemia through self-monitoring of blood glucose, were also classified as GDM (n = 11). All laboratory parameters were measured according to the standard laboratory methods at our certified Department of Medical and Chemical Laboratory Diagnostics (http://www.kimcl.at). Plasma glucose concentrations were measured by

the hexokinase method with a coefficient of variation (CV) of 1.3%. HbA1c was assessed by high performance liquid chromatography, IFCC standardized and DCCT aligned (CV=1.8%). Fifteen clinical prediction models were identified and calculated according to the published literature after performing systematic searches in PubMed, including five 'sum score models',[10-14] nine 'propensity score models',[5,6,15-20] and one decision tree model.[21] A mathematical description of the prediction models is provided in the supplemental material (Table S1). Three additional models were identified but not included in this study as they either used a simple combination of body weight / BMI and maternal age[22,23] or included predictors or biomarkers which are not available in clinical routine, such as tissue plasminogen activator inhibitor or adipokines.[24] Moreover, the full prediction rule was not provided in two of these studies.[22,24]

The study was approved by the Ethics Committee of the Medical University of Vienna (protocol number 1937/2015) and performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants. Reporting of the study conforms to the broad EQUATOR guidelines.[25]

2.2 | Statistical analysis

For statistical analysis, we performed complete case analysis. Women with known glycaemic status in late pregnancy but with variables missing were still included in the analysis. Prediction models containing the missing variables were not calculated for those women.

Continuous variables were summarized by mean ±standard deviation or as median and interquartile ranges (IQR) where appropriate and compared by Welch's t test or the Wilcoxon rank sum test, respectively. Categorical variables were summarized by counts and percentages and compared by Pearson's chi-squared test. Discrimination (ie the ability of a model to separate pregnant women with disease from those without disease) was assessed by receiver operating characteristics curve (ROC) analysis. Calibration (ie the agreement between predicted vs. observed probability of having the disease) was assessed for propensity score models before and after recalibration by use of calibration plots.[26] In addition, we used decision curve analysis to assess the net benefit of the four prediction models showing the best discrimination as compared to default strategies (ie treating all patients or no patients as having the disease).[27] The net benefit is expressed as the net proportion of true positive cases: net benefit = true positive / n – false positive / $n \times w$ where n is the total number of cases and w is a relative weight equal to the odds of a threshold probability. A net benefit of 0.2 means that 20% of the study participants would be correctly classified as true positive cases at a respective cut-off probability of a

model.[26] Finally, random decision forests with ntree = 10^4 were created by the conditional inference framework (cforest) to derive measures of variable importance. Variable importance is calculated as the average difference in predictive accuracy before and after random permutation of the values of a predictor variable over all (ie ten thousand) trees.[28,29] Statistical analysis was performed with R (version 3.5.3) and contributing packages (especially 'pROC', 'rms', 'party' and 'rmda').[30] With a sample size of 1132 patients (239 cases with GDM), we are able to identify an area under the ROC of at least 57.4% compared to the noninformative ROC-AUC value of 50% with a power of 95% and a two-sided α-error probability of 5%. When the total sample size is 1132, where group 1 has a sample size of 239 and group 2 has a sample size of 893, a two-sided 95% confidence level with an area under the ROC equal to 0.65/0.7/0.75 produces a confidence interval width of 0.041/0.040/0.038. A two-sided P-value of ≤.05 was considered statistically significant.

3 | RESULTS

3.1 | Characteristics of the study cohort

Among a total of 1,132 women included in this study, n = 239 (21.1%) developed GDM, whereas n = 893 (78.9%) remained normal glucose tolerant and did not develop GDM (NGT). A comparison of baseline characteristics of both groups is provided in Table 1. Patients who developed hyperglycaemia were older, had significantly higher BMI, body weight and blood pressure already at the beginning of pregnancy and were characterized by higher HbA1c, fasting plasma glucose and triglyceride concentrations. History of GDM in previous pregnancies as well as family history of type 2 diabetes and non-Caucasian ethnicity were more often observed in women who developed GDM. Moreover, history of LGA delivery was more often observed in women who developed GDM.

3.2 | Discrimination of the investigated prediction models

A brief summary of the prediction models analysed in this study including predictor variables and measures of discrimination is presented in Table 2. The areas under the ROC curve (ROC-AUC) ranged between 60.7% and 76.9%, corresponding to a moderate to fair accuracy. In general, 'propensity score' models (ie prognostic models that calculated the probability for GDM on a continuous scale)[5,6,15-20] showed improved predictive performance as compared to 'sum score' models[10-14] and the 'decision tree' model,[21] whereby especially 'propensity score' models showed improved discrimination as compared to maternal age (ROC-AUC:

	n	NGT (n = 893)	n	GDM (n = 239)	P-value
Age (years)	893	31.4 ± 5.8	239	32.8 ± 5.7	<.001
Parity (≥1)	893	541 (60.6)	239	165 (69.0)	.017
GDM in previous pregnancies	893	52 (5.8)	239	68 (28.5)	<.001
Ethnicity (non-Caucasian)	893	184 (20.6)	239	67 (28.0)	.014
Height (cm)	893	165 ± 6.8	239	164 ± 6.8	.008
Weight, before pregnancy (kg)	893	66.1 ± 14.7	239	72.5 ± 16.3	<.001
Weight, current (kg)	893	67.5 ± 14.6	239	74.7 ± 16.4	<.001
BMI, before pregnancy (kg/m^2)	893	24.3 ± 5.2	239	27.1 ± 5.7	<.001
BMI, current (kg/m^2)	893	24.8 ± 5.1	239	27.9 ± 5.7	<.001
Family history (1st degree)	893	214 (24.0)	239	89 (37.2)	<.001
Family history (1st and 2nd degree)	893	386 (43.2)	239	143 (59.8)	<.001
Glucosuria (>40 mg/dL)	893	9 (1.0)	239	3 (1.3)	.740
RRS (mmHG)	892	118 ± 12.8	239	121 ± 11.8	.005
RRD (mmHG)	892	76 ± 10.0	239	79 ± 9.5	<.001
Prior LGA delivery (>90 perc)	893	84 (9.4)	239	38 (15.9)	.004
Prior macrosomia (>4000 g)	893	55 (6.2)	239	27 (11.3)	.006
Preconception dyslipidaemia	891	18 (2.0)	238	7 (2.9)	.391
Assisted reproduction	893	87 (9.7)	239	35 (14.6)	.030
Multiple pregnancy	893	107 (12.0)	239	21 (8.8)	.166
Smoking status	893	352 (39.4)	239	86 (36.0)	.333
FPG (mg/dL)	848	80.6 ± 5.8	224	86.0 ± 7.8	<.001
Triglycerides (mg/dL)	848	107 (82-139)	225	130 (99-165)	<.001
HbA1c (%)	857	4.95 ± 0.29	227	5.13 ± 0.30	<.001

Note: Data are mean ± SD or median (IQR) and count (%) for women remaining normal glucose tolerant (NGT) vs. patients developing gestational diabetes (GDM).

Abbreviations: BMI, body mass index; FPG, fasting plasma glucose; HbA1c, glycated haemoglobin A1c; LGA, large for gestational age neonates; RRD, diastolic blood pressure; RRS, systolic blood pressure.

56.6%, 95%CI: 52.6-60.7), pre-gestational BMI (ROC-AUC: 66.0%, 95%CI: 62.1-69.9) or BMI at study entry (ROC-AUC: 67.5%, 95%CI: 63.7-71.3) alone. The results remained comparable in a sensitivity analysis after excluding mothers with multiple pregnancies (n = 128), although discrimination tended to be lower in another sensitivity analysis including only nulliparous women (n = 426), with ROC-AUC statistics ranging between 56.7% and 73.9% (supplemental material Table S2).

3.3 | Calibration of the investigated prediction models

Calibration plots were assessed for nine propensity score models for which the full prediction rule was published. As visualized in Figure 1, two models showed acceptable calibration (Benhalima-1 and Benhalima-2 2020), two models overestimated (Göbl 2012 and Sweeting 2017) and

six models underestimated the risk of GDM (van Leeuwen 2010, Nanda 2011, Savona-Ventura 2013, Gabby-Benziv 2015 and Syngelaki 2015) when calculated as originally published. The agreement between predicted vs. observed probabilities was notably improved after recalibration (Figure 2): Six out of nine models had acceptable to almost perfect calibration (Göbl 2012, Savona-Ventura 2013, Gabby-Benziv 2015, Sweeting 2017, Benhalima-1 and Benhalima-2 2020), whereas three models showed sporadic overestimation (van Leeuwen 2010, Nanda 2011, Syngelaki 2015).

3.4 | Variable importance and net benefit analysis

Random forest analysis revealed highest variable importance scores for history of GDM in previous gestation as well as laboratory parameters like fasting plasma glucose,

TABLE 2 Summary and discrimination performance of clinical risk prediction models evaluated in this study ordered by type of the model and publication year

Author	Included variables	ROC-AUC (%)	95% CI
Naylor 1997[10]	Sum score model: Age; Pre-pregnancy BMI; Ethnic origin	65.5	61.7-69.2
Caliskan 2004[11]	Sum score model: Age; Pre-pregnancy BMI; Prior adverse obstetric outcome; Family history of diabetes; Prior macrosomia	64.5	60.8-68.3
Shirazian 2009[14]	Sum score model: Age; Pre-pregnancy BMI; Family history of diabetes	60.7	56.8-64.7
Phaloprakarn 2009[13]	Sum score model: Age; First trimester BMI; Family history of diabetes; Prior macrosomia; History of ≥2 abortions	67.6	63.9-71.3
Teede 2011[12]	Sum score model: Age; First trimester BMI; Ethnic origin; Family history of diabetes; History of GDM	68.9	65.2-72.7
Pintaudi 2014[21]	Decision tree model: FPG; pre-pregnancy BMI	67.7	63.8-71.6
van Leeuwen 2010[5]	Propensity score model: Ethnic origin; Family history of diabetes; Multiparity; Pre-pregnancy BMI	70.8	67.1-74.5
Nanda 2011[15]	Propensity score model: Age; First trimester BMI; Ethnic origin; History of GDM; Prior macrosomia	72.9	69.1-76.6
Göbl 2012[17]	Propensity score model: History of GDM; Glycosuria; Age; Family history of diabetes; Preconception dyslipidaemia; Ethnic origin; FPG	71.7	67.7-75.6
Savona-Ventura 2013[20]	Propensity score model: FPG; Age; Diastolic blood pressure	65.2	61.0-69.4
Syngelaki 2015[16]	Propensity score model: History of GDM; First trimester weight; Parity; Age; Height; Family history; Ovulation drugs; Ethnic origin; Birth weight	71.5	67.7-75.3
Gabbay-Benziv 2015[19]	Propensity score model: Age; Ethnic origin; History of GDM; Systolic blood pressure; First trimester BMI	71.6	67.9-75.3
Sweeting 2017[18]	Propensity score model: History of GDM; Ethnic origin; Family history of diabetes; Parity; Age; First trimester BMI	71.2	67.5-74.9
Benhalima-1 2020[6]	Propensity score model: Family history of diabetes; History of smoking before pregnancy; Ethnic origin; Age; Height; First trimester BMI; History of GDM	71.7	68.0-75.4
Benhalima-2 2020[6]	Propensity score model: History of GDM; FPG; Height; Triglycerides; Age; Ethnic origin; First trimester weight; Family history of diabetes; HbA1c	76.9	73.2-80.6

Abbreviations: ROC-AUC, area under the receiver operating characteristic curve; FPG, fasting plasma glucose; BMI, body mass index; HbA1c, glycated haemoglobin A1c.

triglycerides and HbA1c (Figure 3). As a consequence, the Benhalima-2 2020 model, which included these variables, showed superior net benefit (ie the net proportion of true positive cases) as compared to other models. The decision curves of the four models with best discrimination (that is the highest ROC-AUC) are provided in Figure 4.
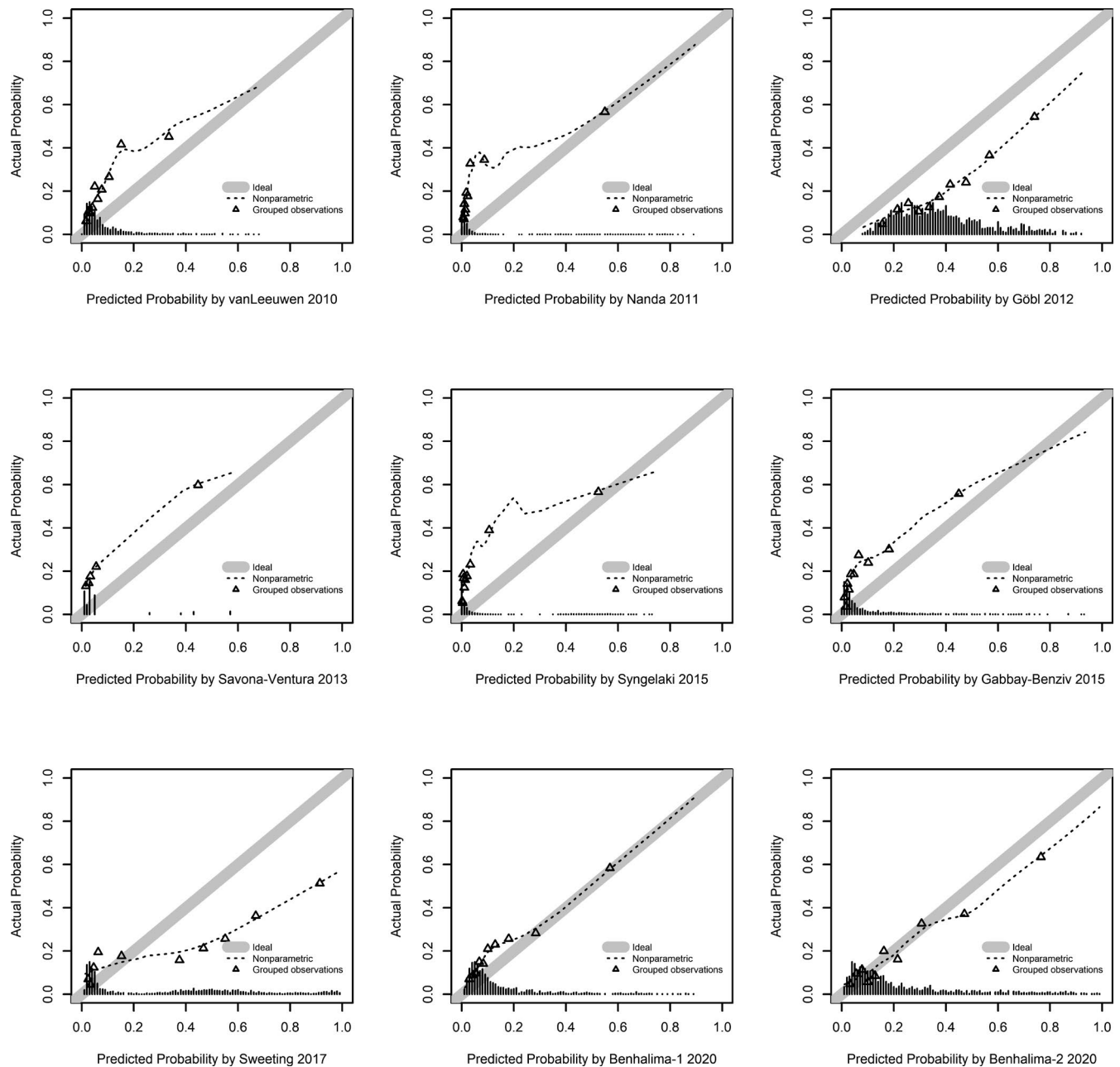
FIGURE 1 Calibration plots of prognostic models (calculated as they were originally published). Good calibration is observed if the dashed calibration line of the model is closely following the ideal calibration line (with an intercept of 0 and a slope of 1) as underlined with grey colour

4 | DISCUSSION

In this prospective validation study, we aimed to investigate the accuracy of low-invasive prognostic models to predict the later development of GDM. A total of fifteen published algorithms were analysed which showed moderate to fair discrimination—this means that their ability to distinguish low from high risk individuals at the beginning of pregnancy was acceptable with areas under the ROC curve up to 76.9%. Although the performance was adequate in terms of discrimination, the agreement between predicted probability and observed risk (ie the calibration) of most models

was poor. A poor calibrated risk estimate, however, is a nonnegligible limitation as it can lead to inaccurate risk predictions and false expectations.[31] Calibration performance (the 'Achilles heel of predictive analytics') is markedly affected by the characteristics of the setting in which a model is developed, including the diagnostic criteria used to define the disease and consequently resulting number of affected cases (low disease incidence in the evaluation cohort can lead to an underestimated risk of the disease, and the opposite for high incidence).[31] This is of particular importance for risk prediction in the field of GDM as the diagnostic algorithms were adapted over time affecting indices of the disease
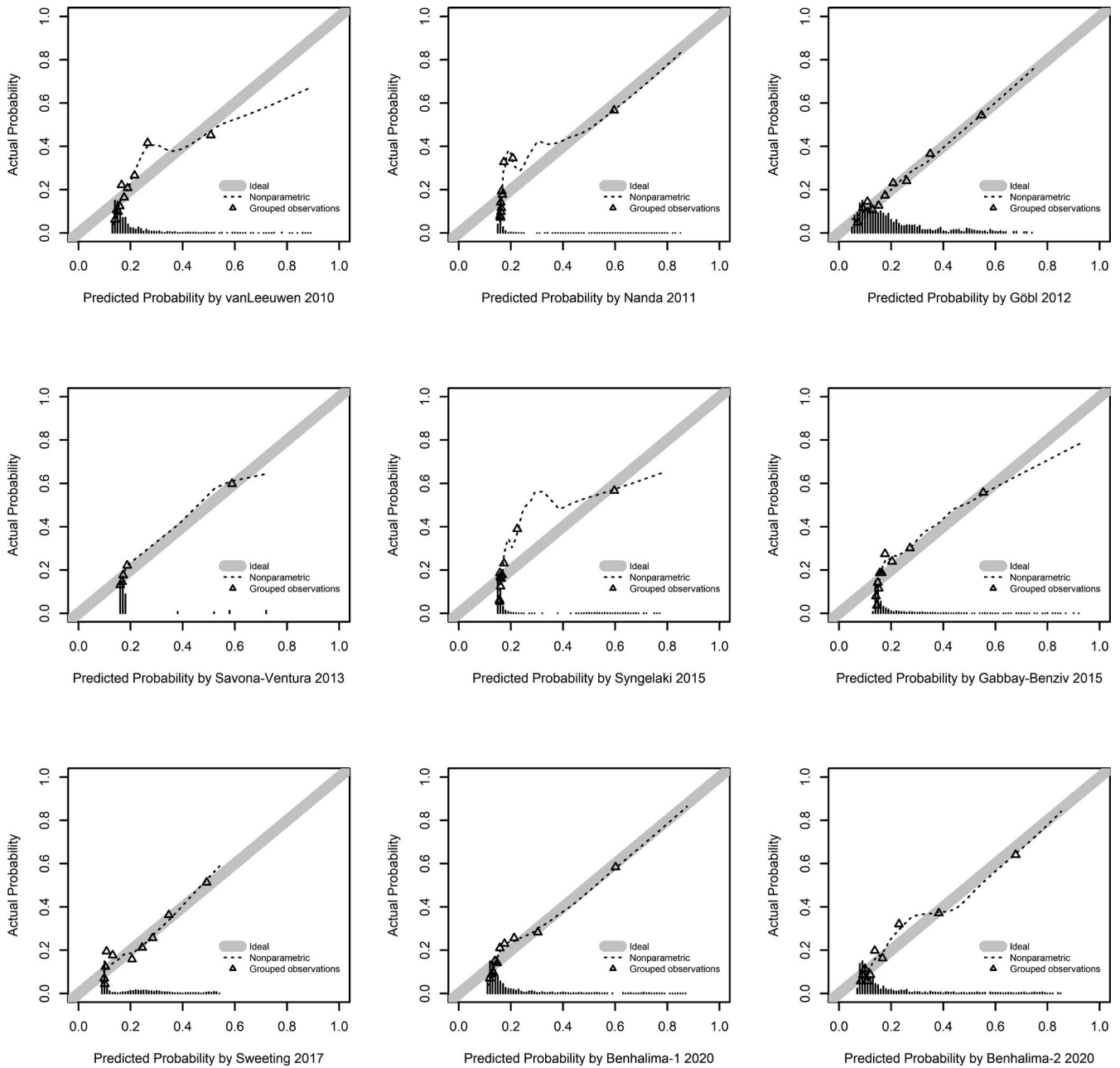
FIGURE 2 Calibration plots of prognostic models (after recalibration). Good calibration is observed if the dashed calibration line of the model is closely following the ideal calibration line (with an intercept of 0 and a slope of 1) as underlined with grey colour

prevalence.[2] Based on the results of the Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) study,[32] the IADPSG proposed novel criteria for the diagnosis of GDM with lower thresholds in 2010, which were later adopted by many health care organizations, including the WHO in 2013, and implemented in many countries.[1,8] Lowering the diagnostic thresholds, however, was associated with identifying more women with moderate hyperglycaemia (ie patients who were previously defined as NGT) and therefore with a higher prevalence of GDM. As a consequence, especially those risk assessment tools that were originally developed based on older criteria (eg van Leeuwen 2010, Nanda 2011, Savona-Ventura 2013,

Syngelaki 2015) showed lower disease incidence (between 2.4% and 8.7%)[5,15,16,20] and hence tended to underestimate the risk for GDM when applied in populations where the most recent IADPSG criteria were considered as is the case in our study. Notably, we observed good calibration for both models most recently proposed by Benhalima et al, who used the recent IADPSG definition to identify women with hyperglycaemia.[6] In our study, two models overestimated the risk for developing GDM (Göbl 2012 and Sweeting 2017)[17,18] and, again, poor calibration can be explained by the respective study settings: The model proposed by Göbl et al was originally developed to estimate the risk for GDM in women with
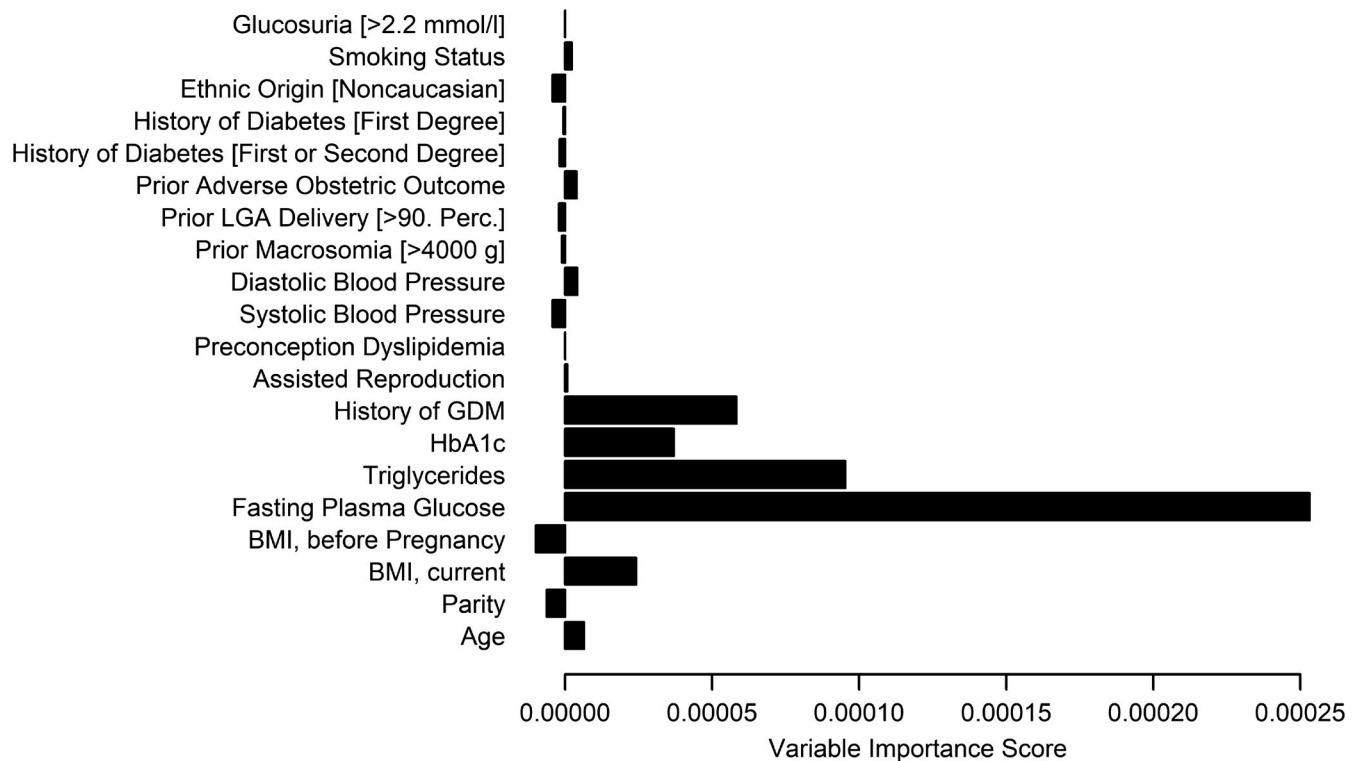
FIGURE 3 Variable importance scores assessed by random forest analysis

normal fasting glucose (< 92mg/dl). Moreover, that study was focussed on women with risk factors and therefore ended up with a markedly high incidence of hyperglycaemia in the evaluation cohort as addressed by the authors.[17] The second study used early OGTTs at the first appointment in patients with risk factors (non-Caucasian ethnicity, age ≥40 years as well as maternal overweight or obesity) which was repeated at 18 to 20 weeks and at 24 to 28 weeks if the previous test was still negative for GDM diagnosis.[18] Repeated OGTT testing before 24 weeks in patients with risk factors is, however, not recommended by other guidelines. Finally, the agreement between observed and predicted risk for GDM was markedly improved by recalibration of the models. Only three recalibrated risk assessment tools showed sporadic overestimation (van Leeuwen 2010, Nanda 2011 and Syngelaki 2015).[5,15,16]

Two further studies systematically evaluated the accuracy of multiple prognostic models to predict the risk for GDM. Lamain-de Ruiter et al assessed twelve risk assessment tools within a well-designed external validation study from the Netherlands and observed ROC-AUC statistics between 67% and 78%.[33] In line with these results, another Dutch study published by Meertens and coworkers reported a discriminative performance of several prediction algorithms between 68% and 75%,[34] which is comparable to our results. The agreement between observed and predicted risk of GDM was improved after recalibration in both studies. In particular, Meertens et al concluded that nearly all models overestimated the risk for GDM when the models were used as

originally published (ie without recalibration). As both Dutch studies used a selective screening approach based on risk factors to identify women with GDM (meaning that only women with risk factors received a diagnostic OGTT where the more stringent WHO 1999 thresholds were applied), both validation cohorts ended up with a low disease incidence of 2.4% (127 affected cases) and 4.9% (181 affected cases), respectively.[33,34] As explained above, different diagnostic criteria and disease incidence in evaluation and validation cohorts could markedly influence model accuracy (especially calibration). This underlines the importance of our study wherein GDM was diagnosed by universal OGTT testing and glucose thresholds in accordance with the most recent guidelines.[1,8]

Our study further extends previous research as we included a detailed examination of variable importance to identify the most prominent risk factors. We found highest variable importance scores for history of GDM in previous pregnancy, as well as elevated routine laboratory parameters, such as fasting plasma glucose, HbA1c and triglycerides. The prognostic value of fasting glucose and HbA1c at early gestation was assessed in previous studies, and discriminative values of approximately 65% for both parameters are reported in the literature.[35-37] In contrast, the prognostic value of maternal lipids is less well investigated, although elevated maternal triglyceride concentrations were found to be associated with adverse pregnancy outcome.[38] Moreover, Bao at al. recently found higher plasma triglycerides in women who developed GDM at early and mid-gestation.[39] It is noteworthy to
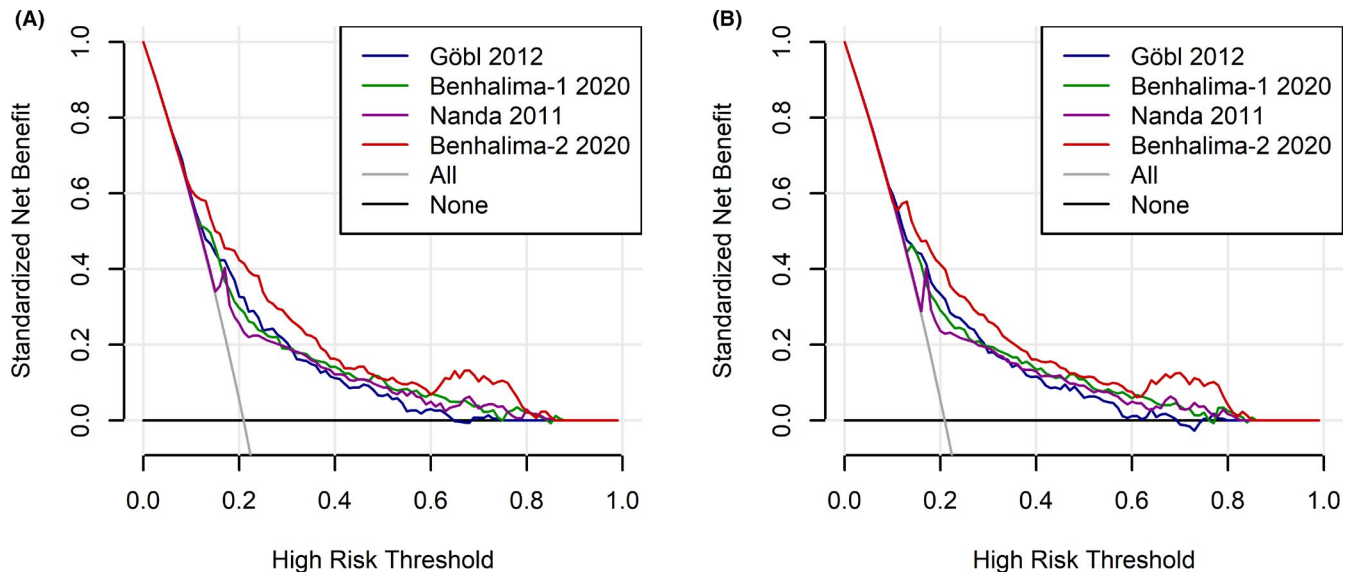
FIGURE 4 Decision curve analysis representing the net benefit (expressed as the net proportion of true positive cases). Decision curves for the four models with best discrimination are shown calculated according to the original publication (A) and after recalibration (B). The solid grey line represents the net benefit when all patients are treated as high risk patients who will develop GDM (All). The solid black line represents the net benefit when none of the patients are considered to develop GDM (None). A model with higher net benefit is preferred

mention that the prediction model that showed the best model accuracy (Benhalima-2 2020) consisted of a combination of anamnestic risk markers in addition to routinely assessable biochemical variables.[6] In addition, the net benefit of this model (defined as the net proportion of true positive cases) across a wide range of threshold probabilities was superior to other algorithms.

In Austria, like in other Central European countries, medical care for pregnant women is provided by gynaecologists and is further supported by general practitioners and midwifes. Patients who develop gestational diabetes are mostly transferred to qualified institutions. This underlines the relevance of clinical prediction models, which can help to identify subgroups with particularly high risk already at early gestation. Moreover, amidst the COVID-19 pandemic we are confronted with challenge of upholding screening standards for GDM while limiting in-person contact for the purpose of reducing the risk of exposure and dissemination of COVID-19. The absence of an adequate screening strategy could result in increased numbers of undiagnosed and untreated GDM cases and consequently GDM related complications, particularly in countries that practice universal screening. Selective screening based on the presence of one or more risk factors has shown to have limited diagnostic accuracy.[4] Under those circumstances, the opportunity arises to utilize risk prediction models in order to improve the existing selective screening algorithms and therefore to effectively reduce the number of inconvenient and unnecessary oral glucose tolerance tests.

Advantages and limitations need to be addressed: The prospective study design as well as the large number of included cases, and especially of women developing GDM, is a clear advantage of this work as risk factors with lower prevalence were captured accurately. Thereby, the number of affected cases is even higher as in other studies in this field.[33,34] Moreover, with our sample size, we were able to detect a prediction model with a ROC-AUC of at least 57.4% as statistically significant. As the predictive accuracy (in terms of discrimination) of several investigated prediction models ranged between 60.7% and 76.9%, we conclude that our study had sufficient power. The single-centre approach could be criticized. However, to our knowledge, this is the first validation study that used the most recent GDM diagnostic recommendations. The disease incidence is comparable with that observed in the HAPO study (ie 18% in the entire HAPO cohort, and between 9.3% and 25.5% in the study centres).[40] Hence, there are no compelling arguments against the validity of our study and the generalizability of our results for Central European settings using the most recent diagnostic criteria. Although ethnic origin was examined in detail in our study, it has to be mentioned that the prevalence of risk factors may vary in different ethnicities and may possibly affect the predictive accuracy of prediction models. Therefore, we recommend further investigations in non-European populations.

There are several concluding remarks: First, model accuracy, and especially calibration, could be influenced by the diagnostic criteria and disease incidence in either evaluation or validation cohorts. This is of particular importance in the field of GDM prediction as diagnostic criteria changed over time and varied between the studies. Second, a detailed assessment of variable importance underlined the additional relevance of routinely available biochemical variables such

as fasting glucose, HbA1c and triglycerides, to further improve clinical prediction models. Third, the model recently proposed by Benhalima et al[6] performed satisfactorily in terms of discrimination and calibration. It can be either used as a basis to develop future models with improved predictive accuracy or tested in clinical studies aiming to reduce the risk for GDM and adverse obstetric outcomes by timely interventions.

CONFLICT OF INTEREST
The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS
CSG conceived the study. Data assessment and patient recruitment was performed by IR, GK, JB, DE and CSG. Calculations and data interpretation were performed by CSG, AT and MM. Statistical analysis was performed by CSG. CSG prepared tables and figures. The manuscript was written by CSG and GK. AT, GYS, CS, JB, IR, PH, DE, EAH and WE critically revised the manuscript. All authors reviewed and edited the final draft of the manuscript. The authors have nothing to disclose. CSG and GK serve as guarantors and accept full responsibility for the work and/or the conduct of the study, had access to the data and controlled the decision to publish.

ORCID
Christian S. Göbl https://orcid.org/0000-0002-3922-7443

REFERENCES
1. World Health Organization. Diagnostic Criteria and Classification of Hyperglycaemia First Detected in Pregnancy. 2013. http://www.who.int/diabetes/publications/Hyperglycaemia_In_Pregnancy/en/
2. Huhn EA, Rossi SW, Hoesli I, Göbl CS. Controversies in screening and diagnostic criteria for gestational diabetes in early and late pregnancy. *Front Endocrinol (Lausanne)*. 2018;9:696. https://doi.org/10.3389/fendo.2018.00696
3. Kjos SL, Buchanan TA. Gestational diabetes mellitus. *N Engl J Med*. 1999;341(23):1749-1756. https://doi.org/10.1056/NEJM199912023412307
4. Griffin ME, Coffey M, Johnson H, et al. Universal vs. risk factor-based screening for gestational diabetes mellitus: detection rates, gestation at diagnosis and outcome. *Diabet Med*. 2000;17(1):26-32.
5. 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*. 2010;117(1):69-75. https://doi.org/10.1111/j.1471-0528.2009.02425.x
6. 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
7. Villar J, Cheikh Ismail L, Victora CG, et al. International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet*. 2014;384(9946):857-868. https://doi.org/10.1016/S0140-6736(14)60932-6
8. International Association of Diabetes and Pregnancy Study Groups Consensus Panel. International Association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care*. 2010;33(3):676-682. https://doi.org/10.2337/dc09-1848
9. Kautzky-Willer A, Harreiter J, Winhofer-Stöckl Y, et al. Gestational diabetes mellitus (Update 2019). *Wien Klin Wochenschr*. 2019;131(Suppl 1):91-102. https://doi.org/10.1007/s00508-018-1419-8
10. Naylor CD, Sermer M, Chen E, Farine D. Selective screening for gestational diabetes mellitus. Toronto Trihospital Gestational Diabetes Project Investigators. *N Engl J Med*. 1997;337(22):1591-1596. https://doi.org/10.1056/NEJM199711273372204
11. Caliskan E, Kayikcioglu F, Oztürk 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. https://doi.org/10.1111/j.0001-6349.2004.00389.x
12. 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. https://doi.org/10.1111/j.1479-828X.2011.01356.x
13. 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. https://doi.org/10.1016/j.ejogrb.2009.04.016
14. 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. https://doi.org/10.1007/s00404-009-1027-y
15. 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. https://doi.org/10.1002/pd.2636
16. 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. https://doi.org/10.1159/000369970
17. Göbl CS, Bozkurt L, Rivic P, et al. A two-step screening algorithm including fasting plasma glucose measurement and a risk estimation model is an accurate strategy for detecting gestational diabetes mellitus. *Diabetologia*. 2012;55(12):3173-3181. https://doi.org/10.1007/s00125-012-2726-7
18. 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. https://doi.org/10.1016/j.diabres.2017.02.036

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

20. Savona-Ventura C, Vassallo J, Marre M, Karamanos BG. MGSD-GDM study group. A composite risk assessment model to screen for gestational diabetes mellitus among Mediterranean women. *Int J Gynaecol Obstet*. 2013;120(3):240-244. https://doi.org/10.1016/j.ijgo.2012.10.016

21. Pintaudi B, Di Vieste G, Corrado F, et al. Improvement of selective screening strategy for gestational diabetes through a more accurate definition of high-risk groups. *Eur J Endocrinol*. 2014;170(1):87-93. https://doi.org/10.1530/EJE-13-0759

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

23. Tran TS, Hirst JE, Do MAT, Morris JM, Jeffery HE. Early prediction of gestational diabetes mellitus in Vietnam: clinical impact of currently recommended diagnostic criteria. *Diabetes Care*. 2013;36(3):618-624. https://doi.org/10.2337/dc12-1418

24. Savvidou M, Nelson SM, Makgoba M, Messow C-M, Sattar N, Nicolaides K. First-trimester prediction of gestational diabetes mellitus: examining the potential of combining maternal characteristics and laboratory measures. *Diabetes*. 2010;59(12):3017-3022. https://doi.org/10.2337/db10-0688

25. Simera I, Moher D, Hoey J, Schulz KF, Altman DG. A catalogue of reporting guidelines for health research. *Eur J Clin Invest*. 2010;40(1):35-53. https://doi.org/10.1111/j.1365-2362.2009.02234.x

26. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

27. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574. https://doi.org/10.1177/0272989X06295361

28. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Stat*. 2006;15(3):651-674. https://doi.org/10.1198/106186006X133933

29. Strobl C, Boulesteix A-L, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. *BMC Bioinform*. 2008;9(1):307. https://doi.org/10.1186/1471-2105-9-307

30. R: The R Project for Statistical Computing. Accessed May 21, 2018. https://www.r-project.org/

31. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17(1):230. https://doi.org/10.1186/s12916-019-1466-7

32. HAPO Study Cooperative Research Group, Metzger BE, Lowe LP, et al. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med*. 2008;358(19):1991-2002. https://doi.org/10.1056/NEJMoa0707943

33. Lamain-de Ruiter M, 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. https://doi.org/10.1136/bmj.i4338

34. 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

35. Zhu W-W, Yang H-X, Wei Y-M, et al. Evaluation of the value of fasting plasma glucose in the first prenatal visit to diagnose gestational diabetes mellitus in China. *Diabetes Care*. 2013;36(3):586-590. https://doi.org/10.2337/dc12-1157

36. Corrado F, D'Anna R, Cannata ML, Interdonato ML, Pintaudi B, Di Benedetto A. Correspondence between first-trimester fasting glycaemia, and oral glucose tolerance test in gestational diabetes diagnosis. *Diabetes Metab*. 2012;38(5):458-461. https://doi.org/10.1016/j.diabet.2012.03.006

37. Hinkle SN, Tsai MY, Rawal S, Albert PS, Zhang C. HbA1c measured in the first trimester of pregnancy and the association with gestational diabetes. *Sci Rep*. 2018;8(1):12249. https://doi.org/10.1038/s41598-018-30833-8

38. Wiznitzer A, Mayer A, Novack V, et al. Association of lipid levels during gestation with preeclampsia and gestational diabetes mellitus: a population-based study. *Am J Obstet Gynecol*. 2009;201(5):482.e1-482.e8. https://doi.org/10.1016/j.ajog.2009.05.032

39. Bao W, Dar S, Zhu Y, et al. Plasma concentrations of lipids during pregnancy and the risk of gestational diabetes mellitus: a longitudinal study. *J Diabetes*. 2018;10(6):487-495. https://doi.org/10.1111/1753-0407.12563

40. Sacks DA, Hadden DR, Maresh M, et al. Frequency of gestational diabetes mellitus at collaborating centers based on IADPSG consensus panel-recommended criteria: the hyperglycemia and adverse pregnancy outcome (HAPO) study. *Diabetes Care*. 2012;35(3):526-528. https://doi.org/10.2337/dc11-1641

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.