



External validation of prognostic models to predict risk of gestational diabetes mellitus in one Dutch cohort: prospective multicentre cohort study

Marije Lamain-de Ruiter,¹ Anneke Kwee,¹ Christiana A Naaktgeboren,² Inge de Groot,³ Inge M Evers,⁴ Floris Groenendaal,⁵ Yolanda R Hering,⁶ Anjoke J M Huisjes,⁷ Cornel Kirpestein,⁸ Wilma M Monincx,⁹ Jacqueline E Siljee,¹⁰ Annawil Van 't Zelfde,¹¹ Charlotte M van Oirschot,¹² Simone A Vankan-Buitelaar,¹³ Mariska A A W Vonk,¹⁴ Therese A Wiegers,¹⁵ Joost J Zwart,¹⁶ Arie Franx,¹ Karel G M Moons,² Maria P H Koster^{1,17}

For numbered affiliations see end of article.

Correspondence to:

M Lamain-de Ruiter
m.deruiter-7@umcutrecht.nl

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2016;354:i4338
<http://dx.doi.org/10.1136/bmj.i4338>

Accepted: 19 July 2016

ABSTRACT

OBJECTIVE

To perform an external validation and direct comparison of published prognostic models for early prediction of the risk of gestational diabetes mellitus, including predictors applicable in the first trimester of pregnancy.

DESIGN

External validation of all published prognostic models in large scale, prospective, multicentre cohort study.

SETTING

31 independent midwifery practices and six hospitals in the Netherlands.

PARTICIPANTS

Women recruited in their first trimester (<14 weeks) of pregnancy between December 2012 and January 2014, at their initial prenatal visit. Women with pre-existing diabetes mellitus of any type were excluded.

MAIN OUTCOME MEASURES

Discrimination of the prognostic models was assessed by the C statistic, and calibration assessed by calibration plots.

RESULTS

3723 women were included for analysis, of whom 181 (4.9%) developed gestational diabetes mellitus in pregnancy. 12 prognostic models for the disorder could be validated in the cohort. C statistics ranged from 0.67 to 0.78. Calibration plots showed that eight of the 12 models were well calibrated.

The four models with the highest C statistics included almost all of the following predictors: maternal age, maternal body mass index, history of gestational diabetes mellitus, ethnicity, and family history of diabetes. Prognostic models had a similar performance in a subgroup of nulliparous women only. Decision curve analysis showed that the use of these four models always had a positive net benefit.

CONCLUSIONS

In this external validation study, most of the published prognostic models for gestational diabetes mellitus show acceptable discrimination and calibration. The four models with the highest discriminative abilities in this study cohort, which also perform well in a subgroup of nulliparous women, are easy models to apply in clinical practice and therefore deserve further evaluation regarding their clinical impact.

Introduction

In the field of obstetrics, the number of publications on prognostic models has more than tripled in the past decade,¹ which reflects an increasing interest in risk based medicine. Risk based medicine aims to provide the most appropriate care to each patient, often guided by outcome risk estimates based on individual patient characteristics, test results, or even genetic information.²

As a result of the obesity pandemic, the incidence of gestational diabetes mellitus, notably occurring in the second or third trimester, is rising and is increasingly contributing to perinatal complications such as macrosomia, shoulder dystocia, caesarean section, and neonatal hypoglycaemia.^{3,4} Moreover, long term sequelae of gestational diabetes mellitus are type 2 diabetes in mothers and obesity in their offspring.^{5,6} Early diagnosis and treatment of gestational diabetes mellitus have been proven to improve pregnancy outcomes.^{7,8} Some guidelines propose a population strategy for diagnosing the disorder⁹⁻¹² (that is, an oral glucose tolerance test) in all pregnant women, whereas others opt for a high risk strategy,¹³ which tests for gestational diabetes mellitus only in women with known risk factors. Both strategies include oral glucose tolerance tests in substantial numbers of women, most of which will lead to negative results, and therefore pose a too high burden to patients as well as health care resources.¹⁴ Accurate

WHAT IS ALREADY KNOWN ON THIS TOPIC

Gestational diabetes mellitus is an increasingly common complication of pregnancy, and pregnancy outcomes can be improved through early screening, diagnosis, and treatment

Many prognostic models estimating the risk of gestational diabetes mellitus have been developed, but an external validation and direct comparison in an independent large cohort of all published models is lacking

WHAT THIS STUDY ADDS

This external validation study shows that in a direct comparison, most published prognostic models for gestational diabetes mellitus in the first trimester have an acceptable discrimination and good calibration

The best performing models in this group can be considered for implementation in routine clinical care

prognostic models for the risk of patients developing gestational diabetes mellitus early in pregnancy could discriminate between high risk and low risk pregnancies, and move towards more tailored care in pregnancy. In particular, this tailored care could result in fewer women undergoing a burdensome diagnostic test (that is, women with a predicted low risk for gestational diabetes mellitus not having to undergo an oral glucose tolerance test).

Several prognostic models for gestational diabetes mellitus have been developed. However, these prognostic models are not commonly used in routine clinical care nor are they recommended by current guidelines. This might be due to the fact that external validation of these prognostic models are scarce,¹⁵⁻¹⁸ let alone that all these models have been directly evaluated and compared on the basis of their predictive accuracy in one independent cohort by independent investigators. To acquire a fair comparison of their predictive accuracy, and thus of their clinical value, it is essential to perform a head-to-head comparison of all published prognostic models in one independent cohort.¹⁹⁻²¹ Thus, the aim of our study was to perform external validation and direct head-to-head comparison of all published first trimester prognostic models for gestational diabetes mellitus, in one independent cohort.

Methods

Study population of external validation cohort

We performed a large prospective multicentre cohort study (the Risk ESTimation for PrEgnancy Complications to provide Tailored care (RESPECT) study) to validate prognostic models for gestational diabetes mellitus. From December 2012 to January 2014, we included pregnant women at their initial prenatal visit (<14 weeks of pregnancy) in 31 independent midwifery practices (primary care) and six hospitals (secondary or tertiary care) in the central region of the Netherlands. We excluded women with any type of pre-existing diabetes mellitus from the cohort. During their pregnancies, participants received routine antenatal care according to Dutch clinical guidelines.

This study was approved by the medical ethics committee of the University Medical Centre Utrecht (protocol no 12-432/C) and written informed consent was obtained from all participants. Results have been reported to conform with the TRIPOD statement.^{22,23}

Predictor assessment

Predictors for gestational diabetes mellitus were all measured in the first trimester at the initial prenatal visit by caregivers or via a self-administered questionnaire. Supplementary appendix A provides detailed information on predictor definition and measurement. We did not report on the distributions of predictors among the original studies that we validated, because it was often missing in the original publications.

Outcome assessment

Gestational diabetes mellitus was diagnosed by a 75 g, two hour, oral glucose tolerance test between 24 and 28

weeks of gestation. According to the World Health Organization 1999 guidelines, which conform with the Dutch national recommendation, the disorder is indicated by the presence of either a fasting glucose level of at least 7.0 mmol/L (126 mg/dL) or a glucose level of at least 7.8 mmol/L (140 mg/dL) after two hours.^{24,25} Women were offered an oral glucose tolerance test if risk factors or any signs of gestational diabetes mellitus were present. Without risk factors or signs, women were not tested and considered as not having gestational diabetes mellitus.

Body mass index in first trimester greater than 30, history of gestational diabetes mellitus, history of macrosomia (birthweight above 95th centile of the Dutch population),²⁶ family history of diabetes mellitus (first degree), non-western ethnicity, history of unexplained intrauterine fetal death, and polycystic ovary syndrome were considered as risk factors for gestational diabetes mellitus. Polyhydramnios and macrosomia were considered as possible signs of the disorder.

For studies validating prognostic models, there is no solid sample size recommendation, but a minimum of 100 patients with events and at least 100 patients without events has been suggested.²⁷

Selection of prognostic models for external validation

In a previous systematic review, we identified 14 published prognostic models for gestational diabetes mellitus that can apply to the first trimester of pregnancy and that only consist of routine measures that are easy to obtain (unpublished data). A short summary of this systematic review is provided in supplementary appendix B.

For proper external validation, it is preferable that the exact definitions of the predictors included in the 14 prognostic models under validation are known as well as how they were measured. Although five of the author groups of the publications of these prognostic models were contacted by email for additional information on intercepts, predictor weights (regression coefficients), and definitions of predictors in the model, none of them responded. Despite this lack of information, it was still possible to include four of these five models in our head-to-head validation study, leaving one model that had to be excluded owing to missing intercept and coefficients.²⁸ Moreover, we excluded one prognostic model from analysis for using maternal abdominal circumference and diagnosis of polycystic ovary syndrome as predictors, which we did not collect in our validation cohort and for which there was also no proxy variable available.²⁹ Thus, a total of 12 prognostic models remained for external validation in the current study.³⁰⁻⁴¹ Supplementary appendix C shows the equations of these 12 prognostic models as applied in our cohort. Table 1³⁰⁻⁴¹ also summarises predictors that were included in the prognostic models. None of the authors of the current external validation study was involved in the development of any of these models.

Table 1 | Summary of predictors for gestational diabetes mellitus per prognostic model³⁰⁻⁴¹

Predictors	Prognostic model												Total No of models
	Caliskan 2004	Eleftheriades 2014	Gabbay-Benziv 2014	Nanda 2011	Naylor 1997	Pintauro 2014	Savona-Ventura 2013	Shirazian 2009	Syngelaki 2011	Teede 2011	Tran 2013	Van Leeuwen 2010	
Maternal age	X	X	X	X	X		X	X	X	X	X		10
Weight		X											1
Body mass index, before pregnancy						X							1
Body mass index	X		X	X	X			X	X	X	X	X	9
Blood pressure			X				X						2
History of gestational diabetes mellitus			X	X						X		X	4
Family history of diabetes mellitus (first or second degree)	X					X		X		X		X	5
History of chronic hypertension									X		X		2
Ethnicity			X	X	X			X	X	X	X	X	6
Parity									X			X	2
Poor obstetric outcome	X												1
History of macrosomia	X			X		X			X				4
Method of conception									X				1
Smoking									X				1
Glucose						X	X						2
Total No of predictors	5	2	5	5	3	4	3	3	8	5	3	5	—

All predictors are measured in the first trimester at the initial prenatal visit unless otherwise specified.

Statistical analysis

Predictor and outcome information was missing for some patients in the validation cohort, and these data were not missing completely at random, as can be derived from table 2. To avoid biased validation of the models, we imputed the missing values using multiple imputation.⁴² All possible predictors and outcomes were used in the imputation model. Ten imputations were performed. Results shown are the results after the multiple imputed data, unless otherwise specified.

Firstly, we applied the original prognostic models—that is, exactly as they were published—to our study cohort when the full prediction rule, including its intercept, was available (supplementary appendix C). Next, to allow for fair comparison of the prognostic models, we performed logistic recalibration by fitting logistic regression models using the linear predictor as the only covariate. This step resulted in an updated calibration slope and intercept.^{43 44}

We assessed discrimination using Harrell's C statistic, which is equivalent to the area under the receiver operating characteristic curve.⁴⁵ It verifies whether participants with a higher predicted risk for gestational diabetes mellitus are indeed more likely to have the disease.

Calibration of the validated original and logistically recalibrated models was assessed by calculation of the predicted probabilities of gestational diabetes mellitus for each individual and comparison of these with their observed outcomes in calibration plots. When a model is well calibrated, the predicted probabilities equal the observed proportions for all groups of predicted probabilities. Thus, when a model is well calibrated, the calibration plot has an intercept of 0 and a slope of 1 and all groups (normally 10 groups) of predicted probabilities fit close to this line. Some calibration plots, however, have fewer than 10 points because it was not possible to split the predicted probabilities into 10 groups. This was the case for models that included only a few categorical variables (eg, sum score models) in which a limited number of predicted probabilities (<10) were possible.

The calibration intercept and slope of the linear prediction after recalibration were used to assess overestimation and underestimation of the models as well as overfitting. A calibration intercept of less than 0 indicates overestimation (the predictions from the original models are too high), whereas an intercept of greater than 0 indicates underestimation. A calibration slope of less than 1 indicates overfitting of the original prognostic models.⁴⁶

A history of gestational diabetes mellitus is an important predictor in most models, but is always scored as negative (=zero) in the prognostic algorithms for nulliparous women, owing to them not having had a previous pregnancy. Therefore, we also reassessed discrimination and calibration of all 12 logistically recalibrated models in a subgroup analysis of nulliparous women to see if the results of the whole population were not merely the result of an excellent

Table 2 | Baseline characteristics of patients in the RESPECT cohort, stratified by variables that were available for imputation

Characteristic	Patients with missing information	Complete cases (n=2603)	Patients with at least one missing value (n=1120)	P	Overall RESPECT cohort (n=3723)*
Age (years)	168 (4.5)	30.9 (4.34)†	30.7 (3.91)†	0.29	30.8 (4.2)†
Body mass index, before pregnancy	46 (1.2)	23.3 (21.2-26.3)†	23.0 (20.9-25.9)†	0.01‡	23.2 (21.1-26.2)†
Body mass index	182 (4.8)	23.8 (21.6-26.9)†	23.4 (21.3-26.3)†	0.01‡	23.7 (21.5-26.7)†
Systolic blood pressure (mm Hg)	65 (1.7)	114 (12)†	115 (12)†	0.48	115 (12)†
Diastolic blood pressure (mm Hg)	64 (1.7)	67 (8)†	67 (8)†	0.09	67 (8)†
Glucose (mmol/L)	171 (4.5)	4.7 (4.3-5.1)†	4.7 (4.4-5.1)†	0.34	4.7 (4.4-5.1)†
Ethnicity					
White		1665 (89.0)	1066 (95.2)		3387 (91.0)
African		17 (0.9)	2 (0.2)		30 (0.8)
Asian	732 (19.7)	30 (1.6)	11 (1.0)	<0.001‡	53 (1.4)
Mixed		44 (2.4)	15 (1.3)		77 (2.1)
Other		115 (4.4)	26 (2.3)		176 (4.7)
Education					
Low		198 (7.6)	52 (4.6)		270 (7.3)
Middle	223 (6.0)	825 (31.7)	362 (32.3)	0.004‡	1273 (34.29)
High		1357 (57.0)	706 (63.0)		2180 (58.6)
Smoking during pregnancy	0	258 (9.9)	73 (6.5)	0.001‡	334 (9.0)
History of chronic hypertension	1 (0.0)	43 (1.7)	14 (1.2)	0.44	57 (1.5)
Family history of diabetes mellitus	1 (0.0)	389 (15.0)	154 (13.8)	0.37	543 (14.6)
Method of conception					
Spontaneous		2396 (93.1)	1033 (92.2)		3429 (92.9)
Ovulation drugs	30 (0.8)	61 (2.4)	38 (3.4)	0.20	99 (2.7)
In vitro fertilisation		82 (3.2)	28 (2.5)		110 (3.0)
Nulliparous	4 (0.0)	1143 (44.0)	509 (45.4)	0.43	1655 (44.5)
History of gestational diabetes mellitus	0	47 (1.8)	12 (1.1)	0.13	59 (1.6)
History of macrosomia (>90th percentile)	0	146 (5.6)	84 (7.5)	0.03‡	230 (6.2)
Recurrent miscarriages (≥2)	4 (0.0)	173 (6.7)	59 (5.3)	0.13	232 (6.2)
History of fetal death	0	58 (2.2)	16 (1.4)	0.14	74 (2.0)
Gestational diabetes mellitus in pregnancy		116 (5.0)	53 (4.7)	0.84	181 (4.9)
Insulin dependent	263 (7.0)	20 (0.8)	13 (1.2)	0.33	33 (0.9)
Gestational age at delivery (days)	342 (9.2)	280 (273-285)†	280 (274-286)†	0.36	280 (273-285)†
Sex (male)	358 (9.6)	1154 (50.7)	569 (52.3)	0.40	1902 (51.1)
Birthweight (g)		3504 (3200-3860)†	3540 (3216-3880)†	0.13	3520 (3190-3880)†
Percentile	372 (10.0)	55 (30-77)†	57 (32-80)†	0.07	55 (30-79)†
>90th percentile		256 (12.0)	140 (13.2)	0.37	494 (13.3)

Data are no (%) unless stated otherwise.

*Data in column includes imputed data for those with missing values.

†Data are mean (standard deviation) or median (interquartile range).

‡Significant at the P<0.05 level.

prediction of gestational diabetes mellitus in multiparous women (that is, women with a known history of the disorder).

Finally, decision curve analysis was performed for the prognostic models with the best discriminative abilities.⁴⁷ Such analyses provide insight into the range of predicted risks for which the model has a higher net benefit than simply either classifying all patients as having the outcome or no (zero) patients as having the outcome. Decision curve analysis can also compare the net benefits of models.

All analyses were done on each of the multiple imputed datasets, and Rubin's rules were used to combine the results into summary estimates. Analyses were performed by the mice and rms packages of R-3.1 for Windows (<http://cran.r-project.org>).

Patient involvement

The HELLP Foundation, a Dutch patient confederation for patients who had a pregnancy complicated by

hypertensive disorders, was involved in defining the research question and the design of the study. Pregnant women were not involved in defining the outcome measures. Focus group interviews with pregnant women participating in the RESPECT cohort were organised to discuss the results and how these results should be implemented into routine care. The final results will be disseminated on the internet through the websites of midwifery practices and regional midwifery collaboration associations.

Results

RESPECT cohort

Of 3723 women included for analysis, 1655 (44%) were nulliparous (fig 1). Table 2 shows the baseline characteristics of these women. Gestational diabetes mellitus was diagnosed in 181 (4.9%) women, 33 (18%) of whom needed insulin for glycaemic control. In the nulliparous subgroup, 71 women (4.3%) developed gestational diabetes mellitus.

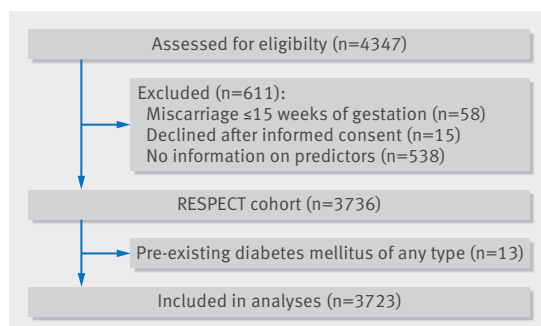


Fig 1 | Flowchart of participants included in validation cohort

Calibration of prognostic models

Three original publications provided the full prediction rule (Gabbay-Benziv 2014, Savona-Ventura 2013, and Van Leeuwen 2010) of which two models showed good calibration, as evidenced both by the calibration plots and the calibration slopes and intercepts (Gabbay-Benziv 2014, Van Leeuwen 2010; fig 2). The model of Savona-Ventura 2013 had a poor calibration, probably partly due to overfitting (slope <1), and tended to overestimate the risk of gestational diabetes mellitus (intercept <0). The calibration plot in fig 2 illustrates how the predicted risks are higher than the observed risks.

However, even though the original intercept was missing, calibration plots were also drawn for each recalibrated model (fig 3). Eight of 12 recalibrated models showed good calibration, with the calibration line closely following the ideal calibration line. The models by Nanda 2011, Pintaudi 2014, Shirazian 2009, and Teede 2011 had sporadic overestimation and underestimation.

Discrimination

C statistics for the original and recalibrated models ranged from 0.67 to 0.78 (table 3). As expected, C statistics of the recalibrated models were slightly lower than those of the original development population. The four models with the highest C statistic (Gabbay-Benziv 2014, Nanda 2011, Teede 2011, and Van Leeuwen 2010) included maternal age, body mass index, history of gestational diabetes mellitus, ethnicity, and family history of diabetes mellitus as predictors. The poorest discriminating models were the models containing the fewest predictors (Eleftheriades 2014, Savona-Ventura 2013, and Tran 2013).

In four prognostic models (Gabbay-Benziv 2014, Nanda 2011, Naylor 1997, and Teede 2011), discrimination for nulliparous women was worse than that for the overall population (table 3). For all other models, the C statistic was higher for nulliparous women only than for all women. Calibration of the prognostic models was also acceptable to good in the nulliparous subgroup (fig 4).

Decision curve analysis

The decision curve analysis results of the four most discriminating models (Gabbay-Benziv 2014, Nanda 2011,

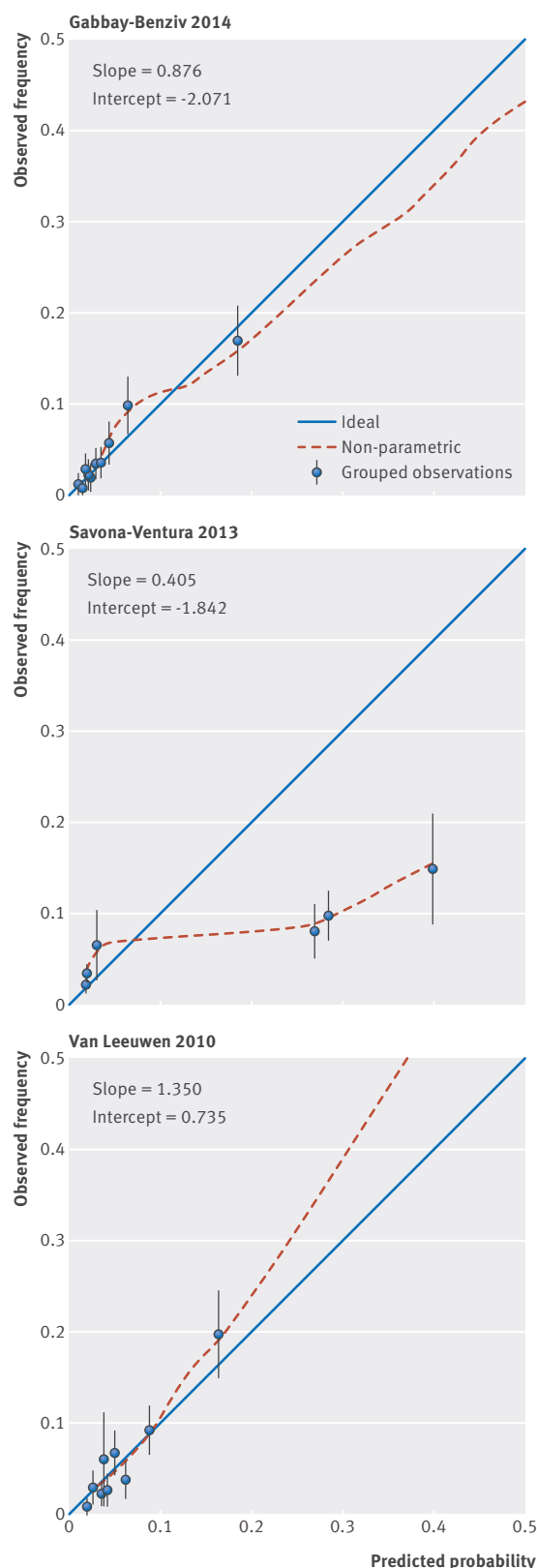


Fig 2 | Calibration plots of original prognostic models to predict risk of gestational diabetes mellitus. In case of perfect calibration, all groups of predicted probabilities fit close to the blue diagonal line, corresponding to an intercept of 0 and a slope of 1 for the calibration plot. Vertical lines in grouped observations represent 95% confidence intervals

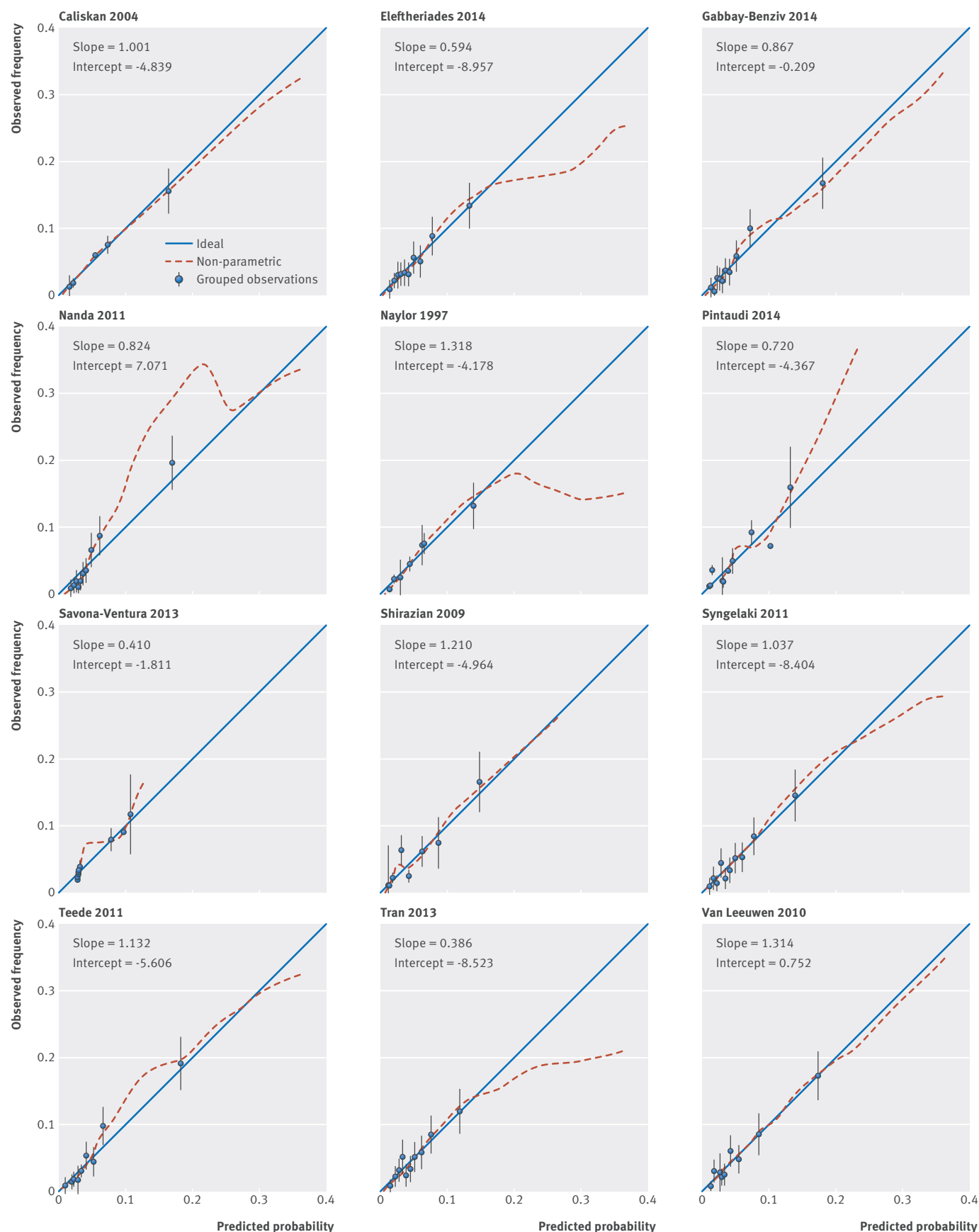


Fig 3 | Calibration plots of recalibrated prognostic models to predict risk of gestational diabetes mellitus. In case of perfect calibration all groups of predicted probabilities fit close to the blue diagonal line, corresponding to an intercept of 0 and a slope of 1 for the calibration plot. Vertical lines in grouped observations represent 95% confidence intervals

Table 3 | Predictive performance of prognostic models for gestational diabetes mellitus in the RESPECT cohort

Prediction model	C statistic (mean (95% CI))		
	Original development models*	Recalibrated models†	Nulliparous women only‡
Caliskan 2004	NR	0.73 (0.69 to 0.76)	0.74 (0.68 to 0.80)
Eleftheriades 2014	0.73 (0.65 to 0.81)	0.70 (0.65 to 0.74)	0.73 (0.66 to 0.79)
Gabbay-Benziv 2014	0.81 (0.77 to 0.87)	0.75 (0.71 to 0.79)	0.72 (0.66 to 0.79)
Nanda 2011	0.79 (0.76 to 0.82)	0.78 (0.74 to 0.82)	0.76 (0.70 to 0.83)
Naylor 1997	0.69 (NR)	0.72 (0.68 to 0.76)	0.71 (0.65 to 0.77)
Pintaudi 2014	NR	0.72 (0.68 to 0.76)	0.73 (0.67 to 0.79)
Savona-Ventura 2013	0.89 (0.86 to 0.91)	0.68 (0.64 to 0.72)	0.72 (0.65 to 0.78)
Shirazian 2009	NR	0.71 (0.67 to 0.75)	0.73 (0.66 to 0.79)
Syngelaki 2011	NR	0.71 (0.66 to 0.75)	0.76 (0.70 to 0.82)
Teede 2011	0.70 (NR)	0.77 (0.73 to 0.81)	0.75 (0.69 to 0.82)
Tran 2013	0.63 (0.60 to 0.65)	0.67 (0.63 to 0.72)	0.69 (0.63 to 0.76)
Van Leeuwen 2010	0.77 (0.69 to 0.85)	0.74 (0.71 to 0.78)	0.74 (0.71 to 0.78)

NR=not reported.

*Shows C statistics as reported in the original publication if available.

†Shows C statistics per model, recalibrated to the RESPECT cohort.

‡Shows C statistics per model when applied to a subgroup of only nulliparous women in the RESPECT cohort.

Teede 2011, and Van Leeuwen 2010) are shown in fig 5. For predicted probability thresholds between 0% and 40%, either prognostic model showed a positive net benefit.

Discussion

Principal findings

A total of 12 first trimester prognostic models for gestational diabetes mellitus were selected by comprehensive review of literature and were compared head-to-head for their predictive accuracy in our population based cohort of 3723 women. Two prognostic models overall (Teede 2011 and Van Leeuwen 2010) had the best performance based on discrimination, calibration, and performance in nulliparous subgroup. Predictors in these particular prognostic models (that is, maternal age, body mass index, ethnicity, parity, history of gestation diabetes mellitus, and history of macrosomia) are easy to measure and widely applicable. Calibration was good for all models and improved by recalibration of the models to our population. Although obstetric history is an important predictor in most models, the prognostic models for gestational diabetes mellitus also performed well in nulliparous women.

Strengths and limitations

This external validation study comprises almost all published, first trimester, prognostic models for gestational diabetes mellitus in one cohort study, allowing for head-to-head comparison of these models. Our study had a large sample size and many cases of gestational diabetes mellitus, used a prospective multicentre approach, and included an unselected population of women from primary care (low risk) as well as secondary or tertiary care (high risk) within a geographically defined area. Additionally, missing data were handled by multiple imputation, which is the most preferable method.⁴⁸

However, some limitations of our study need to be addressed. Firstly, according to Dutch guidelines, a

high risk strategy was adhered. To prevent unnecessary testing in study participants, women without predefined risk factors only underwent an oral glucose tolerance test in case of any symptoms of gestational diabetes mellitus. This strategy could have led to an underestimation of the disorder in low risk women. A study on the performance of similar strategies estimated that 7.3% of diagnoses of gestational diabetes mellitus may have been missed.⁴⁹

However, this possible underestimation is unlikely to have affected the discriminative ability of the validated models, because the C statistic is a rank order insensitive to systematic errors in calibration such as differences in outcome incidence.⁵⁰ Moreover, this potential underestimation is also unlikely to have affected our inferences on the predictive accuracy of the models because we recalibrated the models, which accounts for any differences in overall incidence of gestational diabetes mellitus between the original model development studies and our external validation study. Thus, regardless of any known differences in disease incidence, it is important that prognostic models are recalibrated before they are applied to a new population or when different diagnostic criteria are used.

Secondly, we were not able to include two published prognostic models in our external validation. For one model, information on the prediction rule was not available despite contacting the authors. The other study was published after the start of data collection for our validation cohort, and information on some predictors (such as maternal abdominal circumference and presence of polycystic ovary syndrome) was not collected.

Comparison with other studies

Validation studies on prognostic models for gestational diabetes mellitus are scarce. Our study differs from previous validation studies¹⁵⁻¹⁸ by performing a head-to-head comparison, whereas other studies validated a single prognostic model or only a small selection of the prognostic models for gestational diabetes mellitus. However, our findings are similar to those of these external validation studies, except for the external validation of the Van Leeuwen 2010 model by Lovati and colleagues.¹⁵ In that study, the Van Leeuwen 2010 model yielded a poor C statistic (0.60), by contrast with the other external validation studies (including our current study) that showed C statistics between 0.74 and 0.77.^{17 38} This difference might be due to the case-control study design chosen by Lovati and colleagues, in which it is not possible to adjust for observed outcome frequency.

Clinical implications and conclusions

The use of accurate first trimester prognostic models for gestational diabetes mellitus allows for early and personalised risk stratification in pregnancy. This approach is in contrast to current strategies in which an oral glucose tolerance test is either performed in all women⁹⁻¹² or performed in the presence of any pre-specified risk factors.¹³ Prognostic models have the

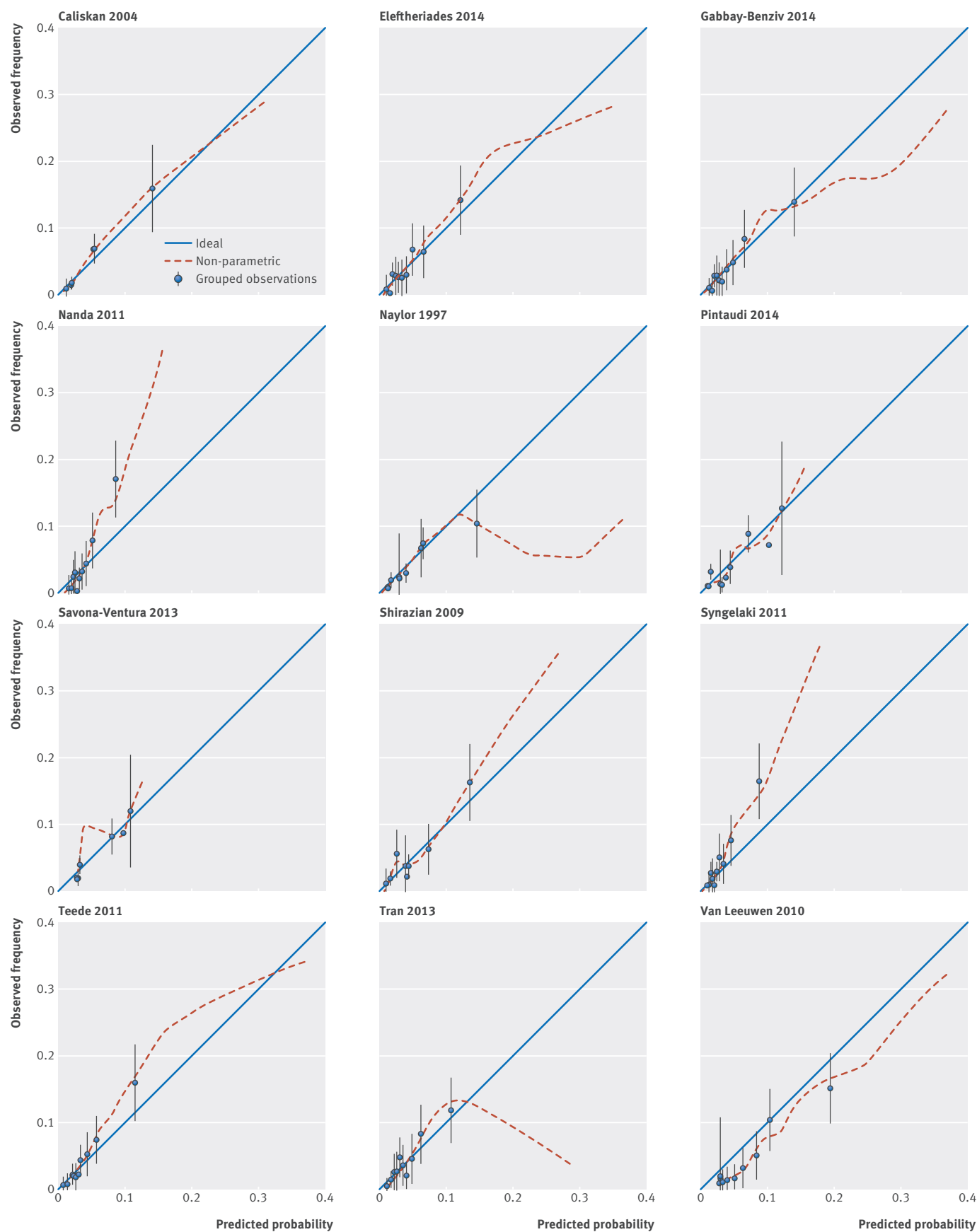


Fig 4 | Calibration plots of recalibrated prognostic models to predict risk of gestational diabetes mellitus, in nulliparous subgroup. In case of perfect calibration all groups of predicted probabilities fit close to the diagonal line, corresponding with an intercept of 0 and a slope of 1 for the calibration plot. Vertical lines in grouped observations represent 95% confidence intervals

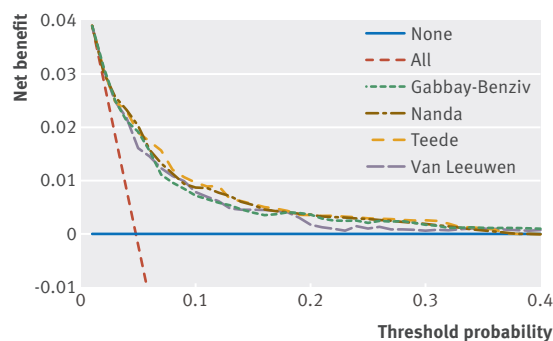


Fig 5 | Decision curve analysis of the four best prognostic models for gestational diabetes mellitus. For the four best prognostic models the net benefit curve is shown. Solid blue line=net benefit when all pregnant women are considered as not having the outcome (gestational diabetes mellitus); red dashed line=net benefit when all pregnant women are considered as having the outcome. The preferred model is the model with the highest net benefit at any given threshold

advantage of being cheap and easy to implement and could avoid the need to perform an oral glucose tolerance test in women with a low risk of developing gestational diabetes mellitus, which relieves both burden and costs.^{9 10} A comparison of performance between the best discriminating prognostic models and current strategies allows an approach that weighs up the pros and cons (eg, missed cases), and will help choose the model to be implemented into clinical practice. The decision on which of the four best models to implement in clinical practice might also depend on population characteristics, availability of predictors, and the incidence of gestational diabetes mellitus, and could therefore be country or region specific.

Implementation of prognostic models for gestational diabetes mellitus early in pregnancy provides room for preventive measures—that is, lifestyle modification interventions such as diet and exercise counselling.^{51 52} Although drug treatments should probably not be the preferred primary intervention to prevent gestational diabetes mellitus, metformin could have a role in the prevention of the disorder in high risk populations.^{53 54} Early prevention, screening, diagnosis, and treatment of gestational diabetes mellitus, when necessary, can and will most likely reduce the rates of caesarean section, neonatal hypoglycaemia and macrosomia, and long term neonatal complications.⁴

In conclusion, most of the 12 previously published prognostic models for gestational diabetes mellitus that have been validated in this study show acceptable to good discrimination and calibration. Four models had C statistics of at least 0.75 (Gabbay-Benziv 2014, Nanda 2011, Teede 2011, and Van Leeuwen 2010). We recommend that these four models be further investigated for implementation in clinical practice. The models by Teede and colleagues and Van Leeuwen and colleagues, with the best overall performance, are easy to apply in clinical practice. The models consist of straightforward predictors: maternal age, body mass index, ethnicity, parity, history of gestational

diabetes mellitus, and history of macrosomia. Once prognostic models for gestational diabetes mellitus are applied in routine clinical care, further research is recommended on the effects on clinical impact, actual development of the disorder, and subsequent pregnancy outcomes.

AUTHOR AFFILIATIONS

¹Department of Obstetrics, Division Woman and Baby, University Medical Centre Utrecht, KE.04.123.1, PO box 85090, 3508 AB, Utrecht, Netherlands

²Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, Netherlands

³Live, Centre for Obstetrics, Tilburg, Netherlands

⁴Department of Obstetrics, Meander Medical Centre, Amersfoort, Netherlands

⁵Department of Neonatology, Division Woman and Baby, University Medical Centre Utrecht, Utrecht, Netherlands

⁶Department of Obstetrics, Zuwe Hofpoort Hospital, Woerden, Netherlands

⁷Department of Obstetrics, Gelre Hospital, Apeldoorn, Netherlands

⁸Department of Obstetrics, Hospital Rivierenland, Tiel, Netherlands

⁹Department of Obstetrics, St Antonius Hospital, Nieuwegein, Netherlands

¹⁰Centre for Infectious Diseases Research, Diagnostics and Screening (IDS), National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands

¹¹Midwifery practice Verloskundigen Amersfoort, Amersfoort, Netherlands

¹²Department of Obstetrics, St Elisabeth Hospital, Tilburg, Netherlands

¹³Midwifery practice GCM, Maarssenbroek, Netherlands

¹⁴Midwifery practice Het Wonder, Houten, Netherlands

¹⁵Netherlands Institute for health services research (NIVEL), Utrecht, Netherlands

¹⁶Department of Obstetrics, Deventer Hospital, Deventer, Netherlands

¹⁷Department of Obstetrics and Gynaecology, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, Netherlands

We thank all the pregnant women who participated in the RESPECT study.

Contributors: MPH, AK, AF, KGMM, and the RESPECT study group (ldG, IME, FG, YRH, AJMH, CK, WMM, JES, AV'tZ, CMvO, SAV-B, MAAWV, TAW, JJZ) had the original idea for the study and were involved in writing the original study protocol. The RESPECT study group and ML-dR were involved in data collection. CAN and ML-dR performed data analysis. ML-dR, CAN, and MPH wrote the first draft of the manuscript, which was subsequently revised by AF, AK, and KGMM. All authors participated in the final approval of the manuscript. MPH and AF are the guarantors of this study. All authors had full access to all of the data (including statistical reports and tables) in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding: This study has been conducted with the support of the Netherlands Organisation for Health Research and Development (project no 50-50200-98-060). The funding source had no role in the design, conduct, analyses, or reporting of the study or in the decision to submit the manuscript for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/doi_disclosure.pdf and declare: support from the Netherlands Organisation for Health Research and Development for the submitted work; no relationships with companies that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: This study was approved by the medical ethics committee of the University Medical Centre Utrecht (protocol no 12-432/C) and written informed consent was obtained from all participants.

Data sharing: Patient level data and full dataset and technical appendix and statistical code are available from the corresponding

author (m.deruiter-7@umcutrecht.nl). Informed consent was not obtained but the presented data are anonymised and the risk of identification is low.

The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>.

- Kleinrouweler CE, Cheong-See FM, Collins GS, et al. Prognostic models in obstetrics: available, but far from applicable. *Am J Obstet Gynecol* 2016;214:79-90.e36. doi:10.1016/j.ajog.2015.06.013.
- James JE. Personalised medicine, disease prevention, and the inverse care law: more harm than benefit? *Eur J Epidemiol* 2014;29:383-90. doi:10.1007/s10654-014-9898-z.
- DeSisto CL, Kim SY, Sharma AJ. Prevalence estimates of gestational diabetes mellitus in the United States, Pregnancy Risk Assessment Monitoring System (PRAMS), 2007-2010. *Prev Chronic Dis* 2014;11:E104. doi:10.5888/pcd11.130415.
- Metzger BE, Lowe LP, Dyer AR, et al. HAPO Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med* 2008;358:1991-2002. doi:10.1056/NEJMoa0707943.
- Bellamy L, Casas J-P, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. *Lancet* 2009;373:1773-9. doi:10.1016/S0140-6736(09)60731-5.
- Clausen TD, Mathiesen ER, Hansen T, et al. High prevalence of type 2 diabetes and pre-diabetes in adult offspring of women with gestational diabetes mellitus or type 1 diabetes: the role of intrauterine hyperglycemia. *Diabetes Care* 2008;31:340-6. doi:10.2337/dc07-1596.
- Alwan N, Tuffnell DJ, West J. Treatments for gestational diabetes. *Cochrane Database Syst Rev* 2009;8:CD003395.
- Viana LV, Gross JL, Azevedo MJ. Dietary intervention in patients with gestational diabetes mellitus: a systematic review and meta-analysis of randomized clinical trials on maternal and newborn outcomes. *Diabetes Care* 2014;37:3345-55. doi:10.2337/dc14-1530.
- American Diabetes Association. Standards of medical care in diabetes—2011. *Diabetes Care* 2011;34(suppl 1):S11-61. doi:10.2337/dc11-S011.
- Nankervis A, McIntyre H, Moses R, et al. ADIPS consensus guidelines for the testing and diagnosis of gestational diabetes mellitus in Australia. Australas Diabetes Pregnancy Soc, 2008: 1-8.
- Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy: a World Health Organization Guideline. *Diabetes Res Clin Pract* 2014;103:341-63. doi:10.1016/j.diabres.2013.10.012.
- Metzger BE, Gabbe SG, Persson B, et al. 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;33:676-82. doi:10.2337/dc10-0719.
- National Institute for Health and Care Excellence. Diabetes in pregnancy. NICE guideline NG3. 2015. <https://www.nice.org.uk/guidance/ng3>.
- Werner EF, Pettker CM, Zuckerwise L, et al. Screening for gestational diabetes mellitus: are the criteria proposed by the international association of the Diabetes and Pregnancy Study Groups cost-effective? *Diabetes Care* 2012;35:529-35. doi:10.2337/dc11-1643.
- 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;100:340-7. doi:10.1016/j.diabres.2013.04.002.
- 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;85:96-101. doi:10.1016/j.diabres.2009.04.025.
- Thériault S, Forest J-C, Massé J, Giguère Y. Validation of early risk-prediction models for gestational diabetes based on clinical characteristics. *Diabetes Res Clin Pract* 2014;103:419-25. doi:10.1016/j.diabres.2013.12.009.
- 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;38:14-21. doi:10.1159/000369970.
- Steyerberg EW, Moons KG, van der Windt DA, et al. PROGRESS Group. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10:e1001381. doi:10.1371/journal.pmed.1001381.
- Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130:515-24. doi:10.7326/0003-4819-130-6-199903160-00016.
- Collins GS, Moons KG. Comparing risk prediction models. *BMJ* 2012;344:e3186. doi:10.1136/bmj.e3186.
- Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55-63. doi:10.7326/M14-0697.
- Correction: transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:600. doi:10.7326/L15-0078-4.
- World Health Organization. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus. 1999. http://apps.who.int/iris/bitstream/10665/66040/1/WHO_NCD_NCS_99.2.pdf.
- Nederlandse Diabetes Federatie (Dutch Diabetes Federation). NDF Zorgstandaard Diabetes 2015. www.zorgstandaarddiabetes.nl/wp-content/uploads/2015/08/NDF-Zorgstandaard-zwangerschapdiabetes-2015.pdf.
- Visser GHA, Eilers PHC, Elferink-Stinkens PM, Merkus HMWM, Wit JM. New Dutch reference curves for birthweight by gestational age. *Early Hum Dev* 2009;85:737-44. doi:10.1016/j.earlhumdev.2009.09.008.
- Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005;58:475-83. doi:10.1016/j.jclinepi.2004.06.017.
- Savvidou M, Nelson SM, Makgoba M, Messow CM, Sattar N, Nicolaides K. First-trimester prediction of gestational diabetes mellitus: examining the potential of combining maternal characteristics and laboratory measures. *Diabetes* 2010;59:3017-22. doi:10.2337/db10-0688.
- Popova PV, Grineva EN, Gerasimov AS, Kravchuk EN, Ryazantseva EM, Shelepova ES. The new combination of risk factors determining a high risk of gestational diabetes mellitus. *Minerva Endocrinol* 2015;40:239-47.
- Caliskan E, Kaykicioglu 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;83:524-30. doi:10.1111/j.0001-6349.2004.00389.x.
- Eleftheriades M, Papastefanou I, Lambrinoudaki I, et al. Elevated placental growth factor concentrations at 11-14 weeks of gestation to predict gestational diabetes mellitus. *Metabolism* 2014;63:1419-25. doi:10.1016/j.metabol.2014.07.016.
- Gabbay-Benziv R, Doyle LE, Blitzer M, Baschat AA. First trimester prediction of maternal glycemic status. *J Perinat Med* 2015;43:283-9. doi:10.1515/jpm-2014-0149.
- 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;31:135-41. doi:10.1002/pd.2636.
- Naylor CD, Sermer M, Chen E, Farine D. Toronto Trihospital Gestational Diabetes Project Investigators. Selective screening for gestational diabetes mellitus. *N Engl J Med* 1997;337:1591-6. doi:10.1056/NEJM199711273372204.
- 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* 2013;170:87-93. doi:10.1530/EJE-13-0759.
- Savona-Ventura C, Vassallo J, Marre M, Karamanos BG. MGS-DGM study group. A composite risk assessment model to screen for gestational diabetes mellitus among Mediterranean women [correction in: *Int J Gynaecol Obstet* 2013;122:88]. *Int J Gynaecol Obstet* 2013;120:240-4. doi:10.1016/j.ijgo.2012.10.016.
- Shirazian N, Emdadi R, Mahboubi M, et al. Screening for gestational diabetes: usefulness of clinical risk factors. *Arch Gynecol Obstet* 2009;280:933-7. doi:10.1007/s00404-009-1027-y.
- 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;30:250-65. doi:10.1159/000328083.
- 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;51:499-504. doi:10.1111/j.1479-828X.2011.01356.x.
- Tran TS, Hirst JE, Do MA, Morris JM, Jeffery HE. Early prediction of gestational diabetes mellitus in Vietnam: clinical impact of currently recommended diagnostic criteria. *Diabetes Care* 2013;36:18-24. doi:10.2337/dc12-1418.
- van Leeuwen M, Opmeer BC, Zweers EJ, et al. Estimating the risk of gestational diabetes mellitus: a clinical prediction model based on patient characteristics and medical history. *BJOG* 2010;117:69-75. doi:10.1111/j.1471-0528.2009.02425.x.

- 42 Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006;59:1087-91. doi:10.1016/j.jclinepi.2006.01.014.
- 43 Janssen KJM, Vergouwe Y, Kalkman CJ, Grobbee DE, Moons KGM. A simple method to adjust clinical prediction models to local circumstances. *Can J Anaesth* 2009;56:194-201. doi:10.1007/s12630-009-9041-x.
- 44 Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med* 2004;23:2567-86. doi:10.1002/sim.1844.
- 45 Harrell F. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer-Verlag, 2001;doi:10.1007/978-1-4757-3462-1.
- 46 Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016;74:167-76. doi:10.1016/j.jclinepi.2015.12.005.
- 47 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565-74. doi:10.1177/0272989X06295361.
- 48 Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014;11:e1001744. doi:10.1371/journal.pmed.1001744.
- 49 Teh WT, Teede HJ, Paul E, Harrison CL, Wallace EM, Allan C. Risk factors for gestational diabetes mellitus: implications for the application of screening guidelines. *Aust N Z J Obstet Gynaecol* 2011;51:26-30. doi:10.1111/j.1479-828X.2011.01292.x.
- 50 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128-38. doi:10.1097/EDE.0b013e3181c30fb2.
- 51 Bain E, Crane M, Tieu J, Han S, Crowther CA, Middleton P. Diet and exercise interventions for preventing gestational diabetes mellitus. *Cochrane Database Syst Rev* 2015;4:CD010443.
- 52 Sanabria-Martínez G, García-Hermoso A, Poyatos-León R, Álvarez-Bueno C, Sánchez-López M, Martínez-Vizcaíno V. Effectiveness of physical activity interventions on preventing gestational diabetes mellitus and excessive maternal weight gain: a meta-analysis. *BJOG* 2015;122:1167-74. doi:10.1111/1471-0528.13429.
- 53 Ainuddin JA, Kazi S, Aftab S, Kamran A. Metformin for preventing gestational diabetes in women with polycystic ovarian syndrome. *J Coll Physicians Surg Pak* 2015;25:237-41.
- 54 Zhuo Z, Wang A, Yu H. Effect of metformin intervention during pregnancy on the gestational diabetes mellitus in women with polycystic ovary syndrome: a systematic review and meta-analysis. *J Diabetes Res* 2014;2014:381231. doi:10.1155/2014/381231.

© BMJ Publishing Group Ltd 2016

Supplementary appendix A: Description of predictors

Supplementary appendix B: Summary of systematic review on prognostic models of gestational diabetes mellitus (unpublished data)

Supplementary appendix C: Full equations for GDM risk prediction models as applied in the validation cohort