Screening for gestational diabetes mellitus

van Leeuwen, M.

Citation for published version (APA):
van Leeuwen, M. (2012). Screening for gestational diabetes mellitus
Chapter 7

Estimating the risk of gestational diabetes mellitus: a clinical prediction model based on patient characteristics and medical history

Marsha van Leeuwen
Brent C. Opmeer
Egbert J.K. Zweers*
Evert van Ballegooie*
Henk G. ter Brugge
Harold W. de Valk
Gerard H.A. Visser
Ben Willem J. Mol

Abstract

OBJECTIVE To develop a clinical prediction rule that can help the clinician to identify women at high and low risk for gestational diabetes mellitus (GDM) early in pregnancy in order to improve the efficiency of GDM screening.

DESIGN We used data from a prospective cohort study to develop the clinical prediction rule.

SETTING The original cohort study was conducted in a university hospital in the Netherlands.

POPULATION Nine hundred and ninety-five consecutive pregnant women underwent screening for GDM.

METHODS Using multiple logistic regression analysis, we constructed a model to estimate the probability of development of GDM from medical history and patient characteristics. Receiver operating characteristics analysis and calibration were used to assess the accuracy of the model.

MAIN OUTCOME MEASURE Development of a clinical prediction rule for GDM. We also evaluated the potential of the prediction rule to improve the efficiency of GDM screening.

RESULTS The probability of the development of GDM could be predicted from ethnicity, family history, history of GDM and body mass index. The model had an area under the receiver operating characteristic curve of 0.77 (95% CI 0.69 - 0.85) and calibration was good (Hosmer and Lemeshow test statistic, $P = 0.25$). If an oral glucose tolerance test was performed in all women with a predicted probability of 2% or more, 43% of all women would be tested and 75% of the women with GDM would be identified.

CONCLUSIONS The use of a clinical prediction model is an accurate method to identify women at increased risk for GDM, and could be used to select women for additional testing for GDM.
Introduction

Gestational diabetes mellitus (GDM) is a common metabolic complication that occurs in 2-9% of all pregnancies.\(^1\) It is well established that GDM is associated with an increased rate of perinatal complications, as well as with a higher maternal risk of development of diabetes mellitus in later life.\(^2,3\) Treatment of women with GDM improves neonatal and maternal outcome significantly.\(^4\) At present, there is no consensus on the optimal strategy for the identification of women with GDM. Several international expert groups recommend the use of clinical risk factors to identify women at risk for GDM.\(^5\) Based on the presence of one or more risk factors, screening or diagnostic testing is offered to these women. The risk factors for GDM reported in the literature are maternal age over 25 years, body mass index (BMI) above 30 kg/m\(^2\), previous macrosomic offspring (>4500 kg), previous GDM, first-degree relative with diabetes and ethnic origin with a high prevalence of diabetes.

Opponents of this selective testing criticise the use of risk factors to select women for screening or diagnostic testing because of limited accuracy.\(^6,8\) The sensitivity and specificity are both considered to be low, leaving women with GDM undiagnosed on the one hand, and leading to unnecessary testing in healthy women on the other. The use of risk factors to identify women at risk for GDM, however, might be effective if their diagnostic value would be specified appropriately in a statistical prediction model. An integrated approach, combining multiple risk indicators, could improve both the accuracy and efficiency of selection of women for additional screening or diagnostic testing. An additional, important benefit of a prediction model based on risk indicators would be that the risk estimation of GDM is assessed early in, or even before, pregnancy. If the risk of GDM can be estimated accurately early in pregnancy, timely interventions during prenatal care could result in maternal and neonatal health benefit.

Although many studies have reported on risk factors for GDM,\(^9-14\) to our knowledge only a few studies have integrated patient characteristics and medical history in a quantitative manner by means of a risk scoring system or a prediction model.\(^15-17\) In this article, we developed a multivariable logistic regression model in which we combined patient characteristics and medical history to predict the occurrence of GDM. We evaluated whether this model was an accurate method to identify women at risk for GDM, and we explored its potential to increase the efficiency of screening for GDM.
Methods

Patients and data
We used data from a prospective cohort study that compared the performance of the 50-g glucose challenge test and the random glucose test as screening tests for GDM. All women with a singleton pregnancy who reported for prenatal care during a period of 2 years were invited to participate in this study. Women with known pre-gestational diabetes mellitus and women who were first seen after 20 weeks of gestation were excluded from the study. The following characteristics were obtained from all participating women at intake: obstetric history (parity, previous miscarriage, history of GDM, history of perinatal death), family history of diabetes mellitus (defined as a first- or second-degree relative with diabetes mellitus type I or II), ethnicity (self-reported), height and weight, age and smoking habits during pregnancy (categorised as smoking or non-smoking). The BMI before pregnancy was calculated as weight (kg)/[height (m)]².

Women underwent random glucose testing as well as a 50-g glucose challenge test. Both tests were performed once between the 24th and 28th week of gestation. If the random plasma glucose was higher than or equal to 6.8 mmol/l, or if the plasma glucose value at 1 hour after administration of 50 g glucose was higher than or equal to 7.8 mmol/l, a 2 hour, 75-g oral glucose tolerance test (OGTT) was performed within 1 week to confirm or rule out the presence of GDM (reference test). The OGTT was performed in the morning after a 12 hour overnight fast and after 3 days of minimal 150-200 g carbohydrate diet. Plasma glucose was determined before and 2 hours after the administration of a 75-g glucose-containing solution. GDM was considered to be present if the 2 hour venous plasma glucose value equalled or exceeded the cut-off value of 7.8 mmol/l, or if the fasting value was >7.0 mmol/l, according to the World Health Organisation (WHO) criteria. If women with two negative screening tests were not tested with the reference test, some women with GDM (with a false-negative screening test result) would remain undetected, consequently generating verification bias. To estimate the proportion of diseased women who were not identified by the screening tests (false-negative fraction), in order to correct for verification bias, we performed an OGTT in a subset of women with two negative screening test results. Subsequently, we used multiple imputation to estimate the results of OGTT in all women in whom no OGTT was performed, based on the results of the two screening tests as well as on patient characteristics. Details of this procedure have been described elsewhere. The original study was performed in two perinatal centres (Isala Clinics in Zwolle and the University Medical Centre in Utrecht). For the development of the prediction model, we used data from one centre (University Medical Centre in Utrecht).
Development of the prediction model

We used multiple logistic regression analysis to develop a statistical prediction model for GDM consisting of medical history and clinical risk indicators. For the development of the model, we evaluated the assumption of linearity in the logistic regression function for the continuous predictor variables age and BMI, using piecewise polynomials (splines) and visual inspection. When the association was found to be non-linear, the variable was transformed to approach linearity. To determine the association between each predictor variable and the occurrence of GDM, we calculated univariable odds ratios (ORs), 95% confidence intervals (95% CIs) and P-values. Subsequently, we performed multivariable logistic regression analysis with a stepwise backwards selection procedure to construct the prediction model. Traditionally, a significance level of 5% in the univariable analysis is required for a variable to enter the multivariable logistic regression model. However, to avoid the erroneous exclusion of a potential relevant predictive variable, we increased the required significance level to enter the model to 30%. A significance level of 20% was applied for a variable to stay in the model. Final model parameters were estimated using the SAS procedure MIANALYZE (multiple imputation procedure), which reflects uncertainty for imputed values using the slightly different estimates of the model parameters of the imputed datasets.

As the performance of prediction models is generally overestimated when applied in clinical practice (optimism), we adjusted the parameter estimates using a shrinkage factor $\lambda$, calculated as $\lambda = (\chi^2 - k)/\chi^2$, where $\chi^2$ is the likelihood ratio test and $k$ is the number of covariates in the model. All model parameters were uniformly shrunken with this shrinkage factor to adjust for optimism. The discriminative performance of the model was assessed by receiver operating characteristic (ROC) curve analysis and calculation of the area under the curve (AUC). An AUC of 0.5 indicates that the scoring system does no better than chance in discriminating between diseased and non-diseased women, whereas a scoring system with perfect discrimination would have an AUC of 1.0. Agreement between the predicted and observed probabilities was evaluated by plotting the mean predicted probabilities in ten risk groups (deciles) as calculated by the model, against the observed proportion of women with GDM in these groups (calibration). The goodness of fit of the model was evaluated with the Hosmer and Lemeshow test statistic.

We evaluated the clinical consequences for different thresholds of the prediction model for subsequent OGTT testing. Finally, we developed a simple scoring system based on the statistical model, in which the probability of GDM can be derived from a nomogram. Data were analysed using SPSS 14.0.1 (SPSS Inc, Chicago, IL, USA), SAS 9.1.3 (SAS Institute Inc, Cary, NC, USA, 2000-2004) and the R computer package (version 2.9.0).
Results

Patients and data
We used data from 995 women who were included in the original cohort study. Random glucose testing was performed in 995 women (100%). The 50-g glucose challenge test was performed in 978 women (98.3%). Thirty-one of the 995 women (3.2%) had an abnormal random glucose value, and 99 of the 978 women (10.1%) had an abnormal result on the 50-g glucose challenge test; 114 women (11.5%) had at least one abnormal test result (random glucose test or 50-g glucose challenge test, or both). In 16 women (1.6%), both test results were suspect for GDM. Of the 114 women with at least one abnormal screening test result, 37 women did not consent to undergo an OGTT. Oral glucose tolerance test was performed in 122 women (12.3%), either because of at least one abnormal screening test result (n = 93) or because the women were part of the subgroup in which OGTT was performed irrespective of two negative screening test results (n=29). Of these 122 women, 22 were diagnosed with GDM and 100 had no GDM. In 873 women, initially no OGTT was performed. The procedure to correct for verification bias indicated that two of these women had GDM, resulting in a total of 24 of 995 women (2.4%) diagnosed with GDM after correction for verification bias.

Development of the clinical prediction model
The assumption of linearity was satisfied in age, but not in BMI. The risk of GDM increased with an increasing BMI between 22 and 30 kg/m², but below 22 kg/m² and above 30 kg/m² the risk of GDM did not alter any further (Figure 1). We therefore performed a
simple transformation for BMI. A BMI below 22 kg/m² was rounded up to 22 kg/m², and a BMI above 30 kg/m² was rounded down to 30 kg/m², for which the association between BMI and the probability of GDM was linear. We planned to evaluate the associations between various ethnicity categories (n = 8) and the occurrence of GDM. However, as the number of women in the various ethnicity categories was too small to demonstrate statistically different associations, in our analyses we only differentiated between Caucasian and non-Caucasian ethnicity.

The results of the univariable and multivariable analyses are summarised in Table 1. The shrinkage coefficient was 0.73, indicating that the performance of the model was overestimated by 27%. The multivariable analysis (after shrinkage) showed that non-Caucasian ethnicity (OR 2.3 (95% CI 1.2 - 4.6)), a family history of diabetes (OR 1.8 (95% CI 0.9 - 3.3)), history of GDM (OR 0.5 (95% CI 0.3 - 1.0) for women without a history of GDM compared to nullipara, and OR 1.6 (95% CI 0.3 - 8.3) for women with a history of GDM compared to nullipara) and BMI (per kg/m²) (OR 1.14 (95% CI 1.04 - 1.25)) increased the risk of GDM. The probability of GDM in our population can be calculated using the formula representing the logistic regression model (after shrinkage): probability of GDM = 1/[1 + exp(−β)], in which β is calculated as [−6.1 + (0.83 × non-Caucasian ethnicity) + (0.57 × positive family history of diabetes mellitus) − (0.67 × multipara without history of GDM) + (0.5 × multipara with history of GDM) + (0.13

<table>
<thead>
<tr>
<th>Table 1. Baseline characteristics of our cohort and results of uni- and multivariable analysis.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Age (per year)</td>
</tr>
<tr>
<td>BMI (per kg/m²)</td>
</tr>
<tr>
<td>Ethnicity non-Caucasian</td>
</tr>
<tr>
<td>Family history of diabetes</td>
</tr>
<tr>
<td>Smoking</td>
</tr>
<tr>
<td>Previous miscarriage</td>
</tr>
<tr>
<td>History of GDM</td>
</tr>
<tr>
<td>Nullipara</td>
</tr>
<tr>
<td>Multipara without history of GDM</td>
</tr>
<tr>
<td>Multipara with history of GDM</td>
</tr>
<tr>
<td>History of perinatal death</td>
</tr>
<tr>
<td>Nullipara</td>
</tr>
<tr>
<td>Multipara without perinatal death</td>
</tr>
<tr>
<td>Multipara with perinatal death</td>
</tr>
</tbody>
</table>

BMI, body mass index; CI, confidence interval; GDM, gestational diabetes mellitus; OR, odds ratio. 95% CI = 95% confidence interval; P = p value. Values represent the number of women (%) unless otherwise indicated. ORs in the multivariable analysis are corrected for over fit with the shrinkage procedure.
The mean AUC of the ROC curves from the ten multiple imputed datasets was 0.77 (95% CI 0.69-0.85), demonstrating a reasonable capacity to discriminate between women with and without GDM. The P-value for the Hosmer and Lemeshow test statistic was 0.25, indicating adequate agreement between the mean predicted probabilities and the observed probability of GDM (calibration).

Clinical consequences
The clinical consequences of several strategies based on the use of the prediction model in our sample are summarised in Table 2. If we applied a predicted probability of 4% as a threshold to consider women ‘at risk’ for GDM, 12.5% of all women would have to undergo OGTT. The positive predictive value of the model in our sample was 8.9% (95% CI 5.5 - 12.5%). The negative predictive value was 98.5% (95% CI 98.0 - 99.0%). If the threshold to proceed to diagnostic testing was set at 2%, 43% of all women would be subjected to OGTT. The positive and negative predictive values were 4.2% (95% CI 3.1 - 4.9%) and 98.9% (95% CI 98.1 - 99.5%), respectively. As one woman with GDM in our cohort was amongst the group of women with the lowest predicted values, there was no predicted probability below which we could ascertain that all women with GDM would be identified when using the prediction model as a screening tool.

Discussion
In this study, we evaluated the use of risk indicators to develop a statistical prediction model for GDM. We found that ethnicity, family history of diabetes mellitus, history of GDM and BMI were independent predictors of GDM in a large cohort of pregnant Dutch women. The use of these simple risk indicators that are easily available from the medical history and demographic characteristics might facilitate the process of screening for GDM. Many studies have been performed to identify risk indicators for GDM. Only a few studies have summarised their results in a prediction model or scoring system in order to provide an estimation of the risk of GDM for every woman individually.16,17

Table 2. Clinical consequences of various strategies.

<table>
<thead>
<tr>
<th>Screening strategy</th>
<th>Women tested with OGTT (n (%))</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>OGTT to diagnose 1 case of GDM (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universal testing with OGTT</td>
<td>995 (100)</td>
<td>100</td>
<td>100</td>
<td>42 (995 / 24)</td>
</tr>
<tr>
<td>OGTT if probability of GDM ≥ 2.0%</td>
<td>428 (43.0)</td>
<td>75.0</td>
<td>57.8</td>
<td>24 (428 / 18)</td>
</tr>
<tr>
<td>OGTT if probability of GDM ≥ 4.0%</td>
<td>124 (12.5)</td>
<td>45.8</td>
<td>88.4</td>
<td>11 (124 / 11)</td>
</tr>
</tbody>
</table>

GDM = gestational diabetes mellitus. OGTT = 75-g oral glucose tolerance test. The number of women in the sample was 995. The number of women with GDM was 24.
The accuracy measures of these scoring systems are summarised in Table 3. Caliskan et al.\textsuperscript{15} developed a scoring system to differentiate between women at low and high risk for GDM. They identified the following risk factors for GDM from a retrospective case-control study in a Turkish population: maternal age $\geq$ 25 years, BMI $\geq$ 25 kg/m\textsuperscript{2}, first-degree relative with diabetes mellitus, previous macrosomic offspring and a previous adverse pregnancy outcome. A score of one point was assigned for the presence of each of the variables. The performance of the scoring system was evaluated in a prospective cohort study, which showed that the prevalence of GDM increased with an increasing score. Selective screening of women with a score $\geq$ 1 decreased the number of screening tests and OGTTs, and all women with GDM were identified. The most important difference between the scoring system used by Caliskan et al. and our prediction model is that, in the scoring system used by Caliskan et al., all predictors were rated equally, ignoring the magnitude of the various risk factors. By quantifying our findings in a statistical prediction model, we can account for the extent to which the individual risk indicators contribute to the risk of GDM. Although the risk scoring system used by Caliskan et al. might have slightly higher accuracy measures, the scoring system was developed using data from women of Turkish origin only, thereby ignoring the influence of ethnicity on the development of GDM, and decreasing the applicability of the scoring system to other populations. In the present study, we accounted for the influence of ethnicity, making our prediction model more applicable to the general population of pregnant women. The slightly lower accuracy measures may also be explained by the more robust methods used in our study for model development.

Another clinical scoring system was developed by Naylor et al.\textsuperscript{16} A clinical scoring system based on age, BMI and ethnicity was developed from a prospective cohort study. The selective screening strategy based on this scoring system showed that the number of screening and diagnostic tests could be decreased for comparable detection rates.

A limitation of our study is that, in the original study, the decision to perform diagnostic testing depended on the results of the two screening tests performed. GDM status was verified in women with a positive screening test result, whereas women with two negative screening tests, in principle, remained unverified. If it is assumed that GDM is absent in unverified women, verification bias would occur because incidental cases of GDM would be missed in this group. We accounted for the problem of verification bias by performing verification with OGTT in a subset of consecutive women with two negative screening tests to estimate the proportion of false-negative test results and multiple imputation of the unverified OGTT results. With this technique, missing OGTT measurements were estimated on the basis of the results of the screening tests as well as patient characteristics. Imputation is a frequently applied technique to deal with missing data, including those resulting from incomplete verification, and is preferred.
over complete case analysis. By using multiple imputation instead of single imputation, the uncertainty regarding imputed values is statistically incorporated, resulting in more accurate confidence intervals.

An important benefit of our statistical model based on patient characteristics and medical history is that the risk estimation of GDM can be performed as early as at the start of pregnancy. From early pregnancy, prenatal care can be provided on the basis of the individual risk profile of GDM. Throughout the whole pregnancy the monitoring and testing of women for GDM can be performed according to the need of the individual patient. In our sample, we detected 75% of women with GDM using this statistical model as a screening tool, with a threshold value of 2% as the predicted probability above which diagnostic testing (OGTT) was performed. This threshold value of 2% was chosen as an example and is arbitrary. To determine the optimal threshold to proceed to diagnostic testing, more information should be obtained on the feasibility of the model in practice, as well as on the preferences of obstetricians, midwives and women concerning the likelihood of detection of GDM, the inconvenience of diagnostic testing and costs.

To use the prediction model in practice, ideally it should be electronically available to the clinician, preferably in an electronic patient file, where the probability of GDM is
Table 3. Comparison of the scoring systems for the risk of gestational diabetes mellitus (GDM)

| Sensitivity (%)(95% CI) | Specificity (%)(95% CI) | PPV (%)(95% CI) | NPV (%)(95% CI) | LR positive test (95% CI) | LR negative test (95% CI) | % women
OGTT |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100 (97.7 - 100.0)</td>
<td>1.47 (1.15 - 1.47)</td>
<td>0.00 (0.00 - 0.67)</td>
<td>69.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99.2 (97.9 - 99.8)</td>
<td>2.41 (1.66 - 2.73)</td>
<td>0.22 (0.06 - 0.62)</td>
<td>37.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98.3 (97.4 - 99.1)</td>
<td>4.16 (2.28 - 6.02)</td>
<td>0.50 (0.25 - 0.78)</td>
<td>15.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97.8 (97.2 - 98.5)</td>
<td>16.2 (6.15 - 37.74)</td>
<td>0.66 (0.44 - 0.85)</td>
<td>3.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96.9 (96.7 - 97.1)</td>
<td>29.1 (3.07 - 275.9)</td>
<td>0.93 (0.87 - 0.99)</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99.1 (98.0 - 99.6)</td>
<td>1.45 (1.32 - 1.52)</td>
<td>0.20 (0.09 - 0.44)</td>
<td>65.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98.0 (97.2 - 98.7)</td>
<td>1.73 (1.46 - 1.95)</td>
<td>0.44 (0.28 - 0.64)</td>
<td>44.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97.6 (96.9 - 98.2)</td>
<td>2.35 (1.86 - 2.82)</td>
<td>0.54 (0.40 - 0.70)</td>
<td>26.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96.5 (96.1 - 97.0)</td>
<td>5.00 (3.08 - 7.79)</td>
<td>0.80 (0.69 - 0.89)</td>
<td>5.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98.9 (98.1 - 99.5)</td>
<td>1.78 (1.30 - 2.10)</td>
<td>0.43 (0.21 - 0.78)</td>
<td>43.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98.5 (98.0 - 99.0)</td>
<td>3.94 (2.34 - 5.77)</td>
<td>0.61 (0.40 - 0.81)</td>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio; OGTT, 75-g oral glucose tolerance test.

The accuracy measures of the scoring systems of Caliskan et al.15 and Naylor et al.16 were calculated from the original articles. Presentation of the results in the studies by Caliskan et al.15 and Naylor et al.16 differ from presentation of the results in the present study. The percentages and points are not directly comparable.

calculated and presented to the clinician. As the use of electronic patient files does not occur in routine clinical practice in a substantial number of obstetric departments and midwife practices, we transformed our prediction model into a paper scoring system in order to facilitate its use in clinical practice (Figure 2). The statistical model developed facilitates clinicians to estimate the probability of GDM. Further (diagnostic) testing or monitoring of women can be individualised on the basis of this probability. To evaluate

Figure 2. Nomogram to estimate the probability of gestational diabetes mellitus (GDM). BMI, body mass index.
the generalisability and applicability of our findings in other populations and different settings, external validation of the prediction model is required. A prediction model should be externally validated before it can be used in clinical practice. The next step in detecting the optimal diagnostic strategy for GDM could involve combining the risk indicators identified in the present study with the results of screening tests, either in a subsequent strategy or an integrated model. Further research should reveal whether this would contribute to an even more accurate strategy, identifying as many women with GDM as early in pregnancy as possible and, at the same time, performing additional testing in as few women as possible, thereby providing adequate healthcare to women with GDM, minimising inconvenience for pregnant women, and saving time and healthcare costs.

Conclusion

We have developed an accurate clinical prediction model for pregnant women that can estimate the risk of GDM at booking based on patient characteristics. The use of a decision rule based on this prediction model could identify women at risk for GDM early in pregnancy, allowing for timely intervention to improve maternal and neonatal outcome.
References


