Screening for gestational diabetes mellitus
van Leeuwen, M.

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

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 2

Comparison of accuracy measures of two screening tests for gestational diabetes mellitus

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

Diabetes Care 2007; 30(11):2779-84
Abstract

OBJECTIVE To compare accuracy measures of the random glucose test and the 50-g glucose challenge test as screening tests for gestational diabetes mellitus (GDM).

RESEARCH DESIGN AND METHODS In this prospective cohort study pregnant women without pre-existing diabetes mellitus in two perinatal centres in the Netherlands underwent a random glucose test and a 50-g glucose challenge test between 24 and 28 weeks of gestation. If one of the screening tests exceeded predefined threshold values, the 75-g oral glucose tolerance test (OGTT) was performed within one week. Furthermore the OGTT was performed in a random sample of women in whom both screening tests were normal. GDM was considered present if the OGTT (reference test) exceeded predefined threshold values. Receiver operating characteristic analysis was used to evaluate the performance of the two screening tests. The results were corrected for verification bias.

RESULTS We included 1301 women. The OGTT was performed in 322 women. After correction for verification bias, the random glucose test showed an area under the ROC curve (AUC) of 0.69 (95% CI 0.61 - 0.78) whereas the glucose challenge test had an AUC of 0.88 (95% CI 0.83 - 0.93). There was a significant difference in area under the curve of the two tests of 0.19 (95% CI 0.11 - 0.27) in favor of the 50-g glucose challenge test.

CONCLUSIONS In screening for GDM, the 50-g glucose challenge test is more useful than the random glucose test.
Introduction

Gestational diabetes mellitus (GDM) is estimated to occur in 2-9% of all pregnancies.\textsuperscript{1-5} It is defined as carbohydrate intolerance with onset or first recognition during pregnancy and is associated with increased rates of adverse pregnancy outcomes, such as macrosomia, shoulder dystocia, birth related trauma such as fractures and nerve palsies and neonatal hypoglycemia and jaundice. In addition, women with GDM are at substantial higher risk to develop diabetes mellitus in later life.\textsuperscript{1, 6-8} Results from a randomized controlled trial show that treatment of GDM by means of dietary advice, blood glucose monitoring and insulin therapy, if required, reduces the rate of serious perinatal complications without increasing the rate of caesarean delivery.\textsuperscript{1}

Based on these results, identification through screening and subsequent treatment of women with GDM appears beneficial. However, consensus on the optimal policy for screening is lacking. The American Diabetes Association recommends screening based on risk factors for GDM (age > 25 years, obesity, close relative with diabetes mellitus, history of GDM, a previous macrosomic infant or specific ethnicity) followed by the 50-g 1 h oral glucose challenge test as a screening test.\textsuperscript{9-11} Other methods of screening that are regularly used are (repeated) random glucose testing, and fasting glucose measurement. It is indefinite which test is the most accurate in testing women for GDM.

The diversity in screening methods may result in unidentified cases of GDM and preventable neonatal and maternal morbidity. Establishment of an optimal, evidence-based screening policy to detect and treat GDM in a timely fashion could contribute to a reduction of perinatal complications. Two regularly used screening tests in the Netherlands are the random glucose test and the 50-g glucose challenge test. The objective of the present study was to compare these two tests as screening tests for GDM as a first step in determining optimal screening policy in GDM.

Research Design and Methods

In a prospective cohort study, all pregnant women attending the outpatient obstetric departments at the University Medical Centre, Utrecht and the Isala Clinics, Zwolle in the Netherlands during a two-year study period, were invited to participate. Women known with pre-existing diabetes mellitus were excluded from the study, as well as those who had not reported for prenatal care in one of the two participating hospitals before 24 weeks of gestation. Only women who delivered after 28 weeks of gestation were included in the analysis.
Data
At intake the following information was obtained: obstetric history, family history of diabetes mellitus, ethnicity (categorized as Caucasian or non-Caucasian), height, self-reported weight (before pregnancy), age and smoking habits (categorized as smoking or non-smoking). The body mass index (BMI) was calculated as weight in kilograms divided by the square of height in meters. The following data regarding pregnancy and outcome were recorded after delivery: weight gain during pregnancy, treatment with diet or insulin, duration of pregnancy in days, birth weight of the neonate in grams, Apgar score after one and five minutes and arterial and venous pH from the umbilical cord.

In all women, the random glucose test was performed at intake (±12 weeks) and between the 24th and 28th week of gestation. If the random plasma glucose measured between 24 and 28 weeks of gestation was higher than or equal to 6.8 mmol/l, the random glucose test was considered abnormal. If random plasma glucose measurement was not performed between the 24th and 28th week, a random plasma glucose at intake higher than or equal to 6.8 mmol/l was considered indicative for GDM. A 50-g oral glucose challenge test was performed between the 24th and 28th week of gestation. The test was performed irrespective of time of the day and of the last meal. Plasma glucose was measured one hour after administration of a solution containing 50 g of glucose. The predefined cutoff value for an abnormal test result was a 1-h plasma glucose value of 7.8 mmol/l.

If either the random glucose test or the 50-g oral glucose challenge test exceeded the predefined threshold value, a two hour 75-g oral glucose tolerance test (OGTT) was performed within one week to confirm or rule out the presence of GDM (reference test). The OGTT was performed in the morning after a 12 hours overnight fast and three days of minimal 150- to 200-g carbohydrate diet. Plasma glucose was determined before and 2 hours after administration of a 75-g glucose containing solution. GDM was considered present if venous plasma glucose equaled or exceeded the threshold values according to the World Health Organization (WHO) criteria (>7.0 mmol/l after 12 hours overnight fast or ≥7.8 mmol/l at two hours after administration of a 75-g glucose containing solution). Venous plasma glucose concentration in all tests was evaluated via glucose oxidase method (Vitros, Otho-Clinical-Diagnostics, Amersham, UK) in the two perinatal centers.

Verification bias
When a screening test is evaluated against a reference test, ideally all participating patients should undergo both the screening and the reference test. However, in practice, the reference test is seldom performed in all patients, as this test is often more invasive or expensive. If only patients with verified screening test results are used to assess the performance of the screening test, calculated accuracy measures become biased because
patients with verified disease status are often only patients with an abnormal screening test result, and therefore they do not represent a random sample of the population in which the screening test is used. The bias that occurs is called (partial) verification bias.12

In the present study, the reference test was, according to the predefined protocol, not performed in all patients. We used the following procedure to correct for verification bias. The OGTT (reference test) was performed in an arbitrary subset of consecutive patients with two negative screening test results to determine the extent to which cases of GDM were missed by the screening tests. Subsequently, we estimated OGTT measurements in women who were not subjected to an OGTT based on results of the random test and the 50-g glucose challenge test as well as on patient characteristics using multiple logistic regression analysis. In other words, if the result of the OGTT was missing, OGTT values were estimated with multiple regression analysis, using the results of the two screening tests and available patient characteristics. This procedure to handle missing data is called imputation and is a commonly used, adequate technique to correct for verification bias.13,14 By using multiple imputation instead of single imputation (i.e., performing the imputation procedure multiple times instead of just once), uncertainty in the imputed values is reflected by the variation in imputed values across multiple imputed datasets, and thus by appropriately larger standard errors (SEs).15 The multiple imputation procedure was also used to impute incidental missing data on patient characteristics.

Statistical analysis
The distribution of continuous variables is reported as mean ± SD. We constructed two-by-two tables for abnormal and normal test results on the random glucose test and the 50-g glucose challenge test against the OGTT. These tables reflect true-positive, false-positive, true-negative or false-negative test results for both the random glucose test and the 50-g glucose challenge test. Diagnostic accuracy (sensitivity, specificity, predictive values and likelihood ratios) and 95% confidence intervals (CIs) were calculated. Receiver operating characteristic (ROC) analysis was used to evaluate the discriminatory power of the two screening tests. Data were analyzed using SPSS 12.0.1 (SPSS, Chicago, IL) and SAS 9.1.3 (SAS Institute Inc. Cary, NC, 2000-2004).

Results
There were 1305 women included in the study. Four women were excluded from analysis because they delivered before 28 weeks of gestation. Data from 1301 women were used for further analysis. Patient characteristics are presented in Table 1. Thereby, the distribution of patient characteristics within the classification groups of the reference test (OGTT) can be compared.
Figure 1 displays the flow of patients in the study based on the results of the subsequent diagnostic test. Of all 1301 women, at least one test result of the random glucose test was obtained. The random glucose test was performed at intake and between the 24th and the 28th week of gestation in 1169 (89.9%) and 1295 (99.5%) of the 1301 women, respectively. We used the results of the random glucose test obtained at intake for six women (0.5%) in whom the random glucose measurement was not performed between the 24th and the 28th week of gestation. None of these six women had a random glucose test result higher than 6.8 mmol/L. The 50-g oral glucose challenge test was performed in 1281 women (98.5%). There were 37 of 1301 women (2.8%) who had an abnormal random glucose test, whereas 167 of 1281 women (13.0%) had an abnormal 50-g glucose challenge test. There were 184 women (14.1%) with at least one abnormal

<table>
<thead>
<tr>
<th>Table 1. Demographics before correction for verification bias.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>GDM present</strong> ( (N = 46) )</td>
</tr>
<tr>
<td>Age (years) ( 30.8 \pm 4.6 )</td>
</tr>
<tr>
<td>BMI before pregnancy ( \text{kg/m}^2 ) ( 25.6 \pm 4.4 )</td>
</tr>
<tr>
<td>Ethnicity ( \text{Caucasian} ) ( 37 % (82.2) )</td>
</tr>
<tr>
<td>( \text{Non Caucasian} ) ( 8 % (17.8) )</td>
</tr>
<tr>
<td>Family history of diabetes ( \text{Yes} ) ( 13 % (28.9) )</td>
</tr>
<tr>
<td>( \text{No} ) ( 32 % (71.1) )</td>
</tr>
<tr>
<td>Smoking ( \text{Yes} ) ( 8 % (17.4) )</td>
</tr>
<tr>
<td>( \text{No} ) ( 38 % (82.6) )</td>
</tr>
<tr>
<td>Hospital ( \text{Utrecht} ) ( 22 % (47.8) )</td>
</tr>
<tr>
<td>( \text{Zwolle} ) ( 24 % (52.2) )</td>
</tr>
<tr>
<td>Obstetric history 1 ( \text{Previous miscarriage} ) ( 15 % (32.6) )</td>
</tr>
<tr>
<td>( \text{No previous miscarriage} ) ( 31 % (67.4) )</td>
</tr>
<tr>
<td>Obstetric history 2 ( \text{Nullipara} ) ( 19 % (43.2) )</td>
</tr>
<tr>
<td>( \text{Multipara with history of GDM} ) ( 2 % (4.5) )</td>
</tr>
<tr>
<td>( \text{Multipara without history of GDM} ) ( 23 % (52.3) )</td>
</tr>
<tr>
<td>Obstetric history 3 ( \text{Nullipara} ) ( 19 % (43.2) )</td>
</tr>
<tr>
<td>( \text{Multipara with perinatal mortality} ) ( 4 % (9.1) )</td>
</tr>
<tr>
<td>( \text{Multipara without perinatal mortality} ) ( 21 % (47.7) )</td>
</tr>
</tbody>
</table>

Data are means \( \pm \) SD or n (%).
test result (random glucose test or 50-g glucose challenge test or both). In 20 women (1.5%) both tests results were suspect for GDM. The OGTT was performed in 322 women (24.8%). This included 146 of the 184 women (79.3%) with an abnormal screening test result and a subgroup of 176 women with two negative screening tests (Figure 1). Initially GDM was diagnosed in 46 women. After correction for verification bias 48 women were diagnosed with GDM (3.7%).

We used multiple imputation of the OGTT values for every patient in whom the OGTT was not performed. This would have been an adequate procedure if the chance of verification of a screening test result depended solely on the result of the screening test. However, we calculated that the chance of verification was not completely independent of factors other than the results of the screening tests. In general, women with a history of GDM or perinatal death, increased BMI and women from the hospital in Zwolle were more likely to be verified, independent of the results of their screening tests. Due to this non-random verification, there was a high prevalence of GDM in women with two negative screening tests who underwent an OGTT. As a result, the prevalence of GDM in the imputed dataset became unrealistically high (up to 15%). In order to obtain imputed data that are in line with the incidence of GDM in the Netherlands (estimated to be approximately 2-4%), we adjusted the imputation procedure by applying the following additional criterion to limit the number of cases classified as having GDM. Based on the same covariates (screening tests and patient characteristics), multiple imputation was

Figures in the diagram represent the number of women with the specific combination of test results before and after correction of verification bias. Figures between parentheses represent the number of women after correction for verification bias.
repeated 100 times and unverified women were only classified as having GDM if they had consistently imputed OGTT values that were indicative for GDM (more than 75%). After this adjusted multiple imputation procedure, the prevalence of GDM in our sample was 3.7%. Only two unverified women were classified as having GDM, whereas in all other women that were unverified, no GDM was assumed.

Table 2 displays results of the comparison of the two screening tests in terms of accuracy measures calculated after correction for verification bias. Comparison of accuracy measures after correction for verification bias resulted in an almost five-times-higher sensitivity in favor of the 50-g glucose challenge test compared to the random glucose test (70.2% (95% CI 57.1 - 83.3) versus 14.6% (95% CI 4.6 - 24.6)). The random glucose test had less false-positive test results and was therefore more specific (97.6% (95% CI 96.6 - 98.5) versus 89.1% (95% CI 87.4 - 90.9)). Positive predictive values for both tests were comparable, as were the negative predictive values. The likelihood ratio of an abnormal test result was larger for the 50-g glucose challenge test than for the random glucose test. The likelihood ratio of a normal test was smaller for the 50-g glucose challenge test. The area under the ROC curve was larger for the 50-g glucose challenge test (0.88 (95% CI 0.83 - 0.93)) than for the random glucose test (0.69 (95% CI 0.61 - 0.78)). There was a significant difference in the areas under the curve of the two tests of 0.19 (95% CI 0.11 - 0.27).

Table 2. Results of the 2x2 table and accuracy measures calculated after correction for verification bias.

<table>
<thead>
<tr>
<th></th>
<th>Random glucose test</th>
<th>1 hour 50-g glucose challenge test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OGGT abnormal</td>
<td>OGGT normal</td>
</tr>
<tr>
<td>Screening test abnormal</td>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>Screening test normal</td>
<td>41</td>
<td>1223</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>1253</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>14.6 (4.6-24.6)</td>
<td></td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>97.6 (96.6-98.5)</td>
<td></td>
</tr>
<tr>
<td>Positive predictive value (%)</td>
<td>18.9 (6.3-31.5)</td>
<td></td>
</tr>
<tr>
<td>Negative predictive value (%)</td>
<td>96.8 (91.0-100.0)</td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio abnormal test result</td>
<td>6.1 (2.8-13.2)</td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio normal test result</td>
<td>0.88 (0.78-0.98)</td>
<td></td>
</tr>
<tr>
<td>Diagnostic Odds Ratio</td>
<td>7.0 (2.9-16.8)</td>
<td></td>
</tr>
<tr>
<td>Area under the curve</td>
<td>0.69 (0.61-0.78)</td>
<td></td>
</tr>
</tbody>
</table>

All accuracy measures are displayed with 95% CIs
Discussion

Evidence for screening for GDM is often inconsistent and difficult to interpret due to various screening methods and thresholds applied internationally. An evidence-based policy could increase the number of identified women with GDM and therefore reduce the number of neonatal and maternal complications by providing adequate monitoring and treatment for these women. For this purpose, the present study compared the random glucose test and the 50-g glucose challenge test as screening tests for GDM. The area under the curve was larger for the 50-g glucose challenge test, indicating that the 50-g glucose challenge test was a better predictor for GDM than the random glucose test.

A potential weakness in present study is the number of missing reference tests, due to which verification bias occurred. Because verification was apparently not performed at random, characteristics other than the screening test results influenced the chance of verification. An intuitive and straightforward procedure to correct for verification bias would be to calculate the ratio diseased / non-diseased from the results of the verified patients stratified by screening test results, and to extrapolate this ratio to the unverified patients. However, this mathematical correction can only be applied if verification of patients is performed completely at random, in other words, if the chance of verification is truly independent of other factors as, for example, patient characteristics. In addition, this results in an adjustment at the sample level. As for individual unverified patients, the disease status according to the reference test remains unknown. To correct for verification bias at the individual level, accounting for factors that influence the chance of verification, imputation techniques can be used to estimate disease status accounting for these factors.

There are several strategies to deal with incomplete data, also within the context of partial verification. As in our study various imputation strategies consistently lead to a considerable higher number of cases, this would consequently imply unrealistically high prevalence rates. We therefore had to apply an additional criterion to limit the number of cases classified as having GDM by means of repeating the multiple imputation procedure for the OGTT 100 times and only classifying women as having GDM if they had consistently imputed values for the OGTT that were indicative for GDM (more than 75 out of 100 times). Further research is required to evaluate which approach is preferred, thereby also accounting for the epidemiological context of the study.

The overall prevalence of GDM in the literature varies from 2.9%. In western countries such as the Netherlands, the prevalence is more often towards 2% than 9%. Hypothetically, the incidence of GDM could be systematically underestimated in the literature (if these estimates have been based solely on selectively verified patients). In that case, we also
underestimated the prevalence of GDM and consequently our approach would have been suboptimal. However, it is not very plausible that for years the incidence of GDM has been underestimated, so application of the described method should have corrected properly for this verification bias.\textsuperscript{18,19}

Results from the present study show that the 50-g glucose challenge test has an almost fivefold higher sensitivity compared to random glucose testing. To our knowledge, these two screening tests have only been equated in the same sample two times before. Mc Elduff et al. found their results in favor of the 50-g challenge test, whereas Mathai et al. found similar sensitivity for both tests and a higher specificity for the random test if both tests were performed in the 26\textsuperscript{th} to 30\textsuperscript{th} week of gestation.\textsuperscript{20,21} A number of studies compared the 50-g glucose challenge test with measurement of fasting glucose. Perucchini et al. found the results in favor of the fasting glucose measurement, whereas Rey et al. showed the 50-g glucose challenge test to be superior.\textsuperscript{22,23} Other studies investigating the test characteristics of the glucose challenge test reported sensitivities ranging from 58 to 80\% for a specificity of around 65\.\textsuperscript{24,25} In these studies, thresholds for an abnormal result of the challenge test ranged from 7.2 to 7.8 mmol/l. In the present study, a predefined cutoff value for an abnormal test result was set at 7.8 mmol/l. If thresholds were set lower than 7.8 mmol/l, sensitivity of the 50-g glucose challenge test would increase, at the expense of a decreased specificity.

The random glucose test is a fast, simple and relatively inexpensive test. Accuracy of random glucose measurement is less frequently studied than the glucose challenge test. Nasrat et al. evaluated random glucose measurement, which revealed a sensitivity of 16\% and a specificity of 96\% using a threshold value of 7.0 mmol/l or 6.4 mmol/l if evaluated within or more than 2 hours postprandial.\textsuperscript{26} Jowett et al. also concluded that random glucose measurement is not sufficiently sensitive for screening on GDM.\textsuperscript{27} Results from the present study are in accordance with results from those two groups, using a threshold value for an abnormal test result of 6.8 mmol/l. As high sensitivity is key to any screening test, random glucose testing is not an accurate method to screen women for GDM, as still five out of six women with GDM would be missed.

**Conclusion**

In conclusion we recommend that despite easy implementation, low costs and relative high specificity, random glucose measurement should not be used as a screening test for GDM. Until superior screening alternatives become available, the 50-g glucose challenge test should be preferred as screening test for GDM.
References


