Screening for gestational diabetes mellitus
van Leeuwen, M.

Citation for published version (APA):
Chapter 3

Accuracy of the random glucose test as screening test for gestational diabetes mellitus

A systematic review

Marsha van Leeuwen
Brent C. Opmeer
Yildirim Yilmaz
Jacqueline Limpens
Mireille J. Serlie
Ben Willem J. Mol

European Journal of Obstetrics & Gynecology and Reproductive Biology.
2011; 154(2):130-5
Abstract

OBJECTIVE Although not formally supported by guidelines, random glucose testing (RGT) is frequently used to screen for gestational diabetes mellitus (GDM). Results on test accuracy are inconclusive. The aim of this study was to systematically review the literature and calculate summary estimates of accuracy measures of RGT as screening test for GDM.

STUDY DESIGN Systematic review to identify studies comparing RGT to oral glucose tolerance testing (75 or 100-g OGTT) before 32 weeks of pregnancy. A systematic search without language restrictions was performed in MEDLINE (1950 to April 2008) and EMBASE (1980 to April 2008). Study selection and data extraction was performed by two independent reviewers. Outcome measures were summary estimates of test accuracy of RGT.

RESULTS Six studies were included, reporting on 3537 women. Due to the small number of studies and heterogeneity, no summary estimates of test accuracy were calculated. Reported sensitivities and specificities of individual studies varied. For 100% sensitivity, specificity was around 40%. For sensitivity of 60% specificity was at most 80%. When specificity approached 100%, sensitivity dropped to 20-30%.

CONCLUSION Available evidence on accuracy of RGT to test for GDM is limited. Based on studies in our systematic review, we consider single random glucose measurement inadequate to screen for GDM.
Introduction

Gestational diabetes mellitus (GDM) is a metabolic complication that occurs in 2-9% of all pregnancies and is associated with increased neonatal and maternal morbidity. Treatment of GDM improves perinatal as well as maternal outcome. Whether screening for GDM will result in reduction of maternal and neonatal morbidity remains to be established. The majority of international diabetes associations however, advocate screening for GDM as desirable. Currently there is no consensus on the optimal approach to screen for GDM. Several international guidelines recommend either a one-step 75-g oral glucose tolerance test (OGTT) approach, or a two-step approach in which a 50-g glucose challenge test is performed, followed by an OGTT in the event of an abnormal test result. Results of the Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) study support the use of the former.1

Although not supported in clinical guidelines, various other tests are used to screen for GDM. One of these is the random glucose test (RGT). A national survey from the UK showed that 52% of the respondents used random glucose measurement to test for GDM. Results from a Dutch survey showed similar results. In 46% of the participants random glucose testing was the most frequently used method to screen for GDM. The RGT is a simple, fast and inexpensive test, which measures plasma glucose at a random point in time, irrespective of the time of the last meal and without any specific preparation. There seem to be only few studies on the accuracy of the RGT as a screening test for GDM. Results of these individual studies suggest that screening with the RGT might lead to considerable false-positive and false-negative test results, although results are not conclusive. The use of inadequate screening methods can result in unidentified cases of GDM and therefore preventable maternal and neonatal morbidity. In addition, it can result in avoidable health care costs due to testing strategies that result in false-positive cases. An accurate evidence-based method for screening could ameliorate the process of diagnosis and management of GDM, resulting in reduction of the rate of serious perinatal complications and maternal morbidity as well as in reduction of healthcare costs.

As high accuracy, especially high sensitivity, is an important prerequisite for screening procedures, the RGT should not be used as screening test for GDM if test accuracy indeed is insufficient, even if the test is simple and inexpensive. If, on the other hand, the accuracy of the RGT is sufficient, more complex screening tests for GDM could be abandoned. The aim of this study was to systematically review the literature and to calculate summary estimates of accuracy measures of the RGT in pregnant women in order to assess its suitability to screen for GDM.
Material and Methods

Literature search
A medical librarian (J.L.) undertook a systematic search in the electronic databases MEDLINE (1950 to April 2008) and EMBASE (1980 to April 2008) to identify studies reporting on the RGT in pregnant women. In accordance with recommendations for Cochrane systematic reviews of diagnostic accuracy we initially searched broadly for the target disease (GDM) and the index test (RGT) using both free-text words and Subject Headings. No methodological filter for diagnostic accuracy studies or any other restriction was applied as this can lead to omission of relevant papers. To find diagnostic accuracy papers that did not mention random glucose test in title and/or abstract we also searched EMBASE for target disease combined with Subject Headings for diagnostic studies. Similar diagnostic index terms are not available for MEDLINE. We systematically inspected reference lists, conducted a “cited reference search” in Web of Science, applied “related articles/find similar feature” in PubMed and Embase, and contacted authors of primary studies for further published trials. We downloaded all references identified into Reference Manager® software version 11.0 (Thomson ISI ResearchSoft, Carlsbad, CA, USA). Duplicate studies were excluded.

Study selection
Two reviewers (M.v.L. and Y.Y.) independently screened titles and abstracts of all retrieved studies. If either reviewer concluded that the article would possibly fulfill eligibility criteria, we obtained the full text publication. Based on the full text manuscripts, the two reviewers selected studies according to predefined criteria. Eligible studies compared the RGT to the 75 or 100-g oral glucose tolerance test (OGTT) (reference test) in pregnant women before 32 weeks of gestation and reported sufficient data to construct a two-by-two table of test performance. Studies that did not report sufficient data to construct a two-by-two by table, but for which data could possibly be obtained from the authors, were also evaluated. Final in-/exclusion decisions were made by comparing results of both reviewers. Disagreement was resolved by consensus or by consulting a third reviewer (B.W.M.).

Data extraction
Data were extracted using a pre-designed piloted data extraction form. We extracted data on study design, sample characteristics and test characteristics, including test accuracy. Data on test accuracy were abstracted as two-by-two tables cross-classifying results of the RGT with results of the OGTT. In case of multiple publications of one study, we used all publications to acquire complete data. The most recent and complete results were included in the analysis. If there were data missing concerning test accuracy, we contacted the corresponding author by email or by letter. Disagreement on data was resolved by discussion and consensus or by consulting a third reviewer (B.W.M.).
Study quality
We evaluated methodological quality of selected studies with QUADAS, a tool for quality assessment of diagnostic accuracy studies.\textsuperscript{17} Included studies were evaluated by two reviewers (M.v.L. and Y.Y.) on 15 items concerning selection, verification, description of tests and of study population.

Analysis
For all included studies we calculated sensitivity and specificity with 95% confidence intervals. To assess heterogeneity of the results, we plotted sensitivity against 1-specificity for all studies in a receiver operating characteristic (ROC) plot. To calculate summary estimates of sensitivity and specificity with 95% confidence intervals, we intended to use a bivariate regression model. With a bivariate regression model summary estimates of sensitivity and specificity can be calculated simultaneously, accounting for the possible correlation between these measures.\textsuperscript{18} However, because of the small number of included studies and because of the clinical heterogeneity of studies that were included we considered meta-analysis not appropriate. Statistical analyses were performed using SPSS 14.0 (SPSS Inc., Chicago, IL, USA).

Diagnosis of GDM
The reference test to diagnose GDM was either the 75-g or 100-g OGTT. Various thresholds for an abnormal OGTT are in use. In the past, impaired glucose tolerance (IGT) was considered to be a condition in between normoglycemia and GDM. Nowadays, the IGT classification is not often used anymore. According to for example the World Health Organization (WHO) or the American Diabetes Association (ADA) criteria women are classified as being normoglycemic or as having GDM.\textsuperscript{6,19} To be able to compare the studies by reference test in the systematic review, original classifications were sometimes abandoned, and women classified as having impaired glucose tolerance in the original article were classified as either being normoglycemic or as having GDM according to currently used criteria.\textsuperscript{6,19}

Results
With the systematic literature search we identified 322 studies. Figure 1 summarizes the process of literature identification and study selection. We selected 27 studies for further reading. Nine authors were contacted for additional data of whom five authors responded. Only one author was able to provide additional data. Eight studies of which no useful data could be obtained were excluded. We excluded 13 other studies for various reasons (Figure 1). The main reason for exclusion was partial or selective verification of the RGT results. Thus, six studies remained for further analysis (Table 1). Table 2 displays study quality as evaluated with the adjusted QUADAS list. Four studies gave a clear...
description of the RGT and of the OGTT. The time between the RGT and the OGTT was < 14 days in three studies, between 14 to 28 days in one study and < 28 days in one study. In one study the time between the tests could not be retrieved. None of the studies met all criteria on the QUADAS list. Post hoc the following items were considered to define a study as high quality: prospective recruitment, consecutive inclusion of all pregnant women with adequate description of inclusion criteria, adequate description of the index test and 100% verification of the index test. None of the studies met all criteria to be labeled as high quality.

Characteristics of included studies are summarized in Table 1. In four studies the RGT was performed only once during pregnancy. In two studies multiple RGTs were performed. Four studies in which the RGT was performed once during pregnancy all were prospective cohort studies with consecutive recruitment of all pregnant women. These studies comprised a total of 2678 women of whom 217 (8.1%) developed
In one study separate thresholds for an abnormal RGT result were set for women who had and who had not eaten within 2 hours of the RGT. Accuracy measures of the RGT were reported for the two thresholds together. Sensitivity reported in the four studies ranged from 15% (95% CI 8 - 25) to 100% (95% CI 75 - 100) depending on the threshold that was applied, with a corresponding specificity of 98% (95% CI 97 - 98) and 37% (95% CI 35 - 37). The study with the best test accuracy had a sensitivity of 64% and a specificity of 80%.

The fifth study that was included was a prospective cohort study, in which the RGT and the OGTT were both performed twice and 749 consecutive women underwent both tests in the first trimester of pregnancy. In women with a normal OGTT result (n = 735), the RGT as well as the OGTT were repeated in the second trimester. Accuracy measures were calculated separately for both trimesters. In the first trimester sensitivity was 71% (95% CI 46 - 88) with corresponding specificity of 80.3% (95% CI 79.8 - 80.6) for a threshold of 5.3 mmol/L. In the second trimester sensitivity was reported to be of 38% (95% CI 14 - 69) with a corresponding specificity of 82.3% (95% CI 82.0 - 82.6) for a threshold of 5.3 mmol/L. The sixth study that was included was a prospective cohort study in which 110 women with risk factors for GDM (e.g. poor obstetric history (not further specified in the original article) and family history of diabetes mellitus) were admitted to the hospital for five venous plasma glucose measurements in 24 hours at 27 to 31 weeks of pregnancy. A 75-g OGTT was performed on the same day. Accuracy measures were calculated for all five RGT measurements. The accuracy measures that we calculated based on information from the article did not match with the accuracy measures reported in the original article. For a threshold of 5.6 mmol/L, the lowest sensitivity of the five measurements that we calculated was 25% (95% CI 18 - 27) and the highest sensitivity was 47% (95% CI 37 - 56) with corresponding specificities of 97% (95% CI 91 - 99) and 74% (95% CI 66 - 81).

Sensitivity and specificity of all studies were plotted in an ROC space (Figure 2). We selected one of the five measurements of the study by Jowett et al. We selected only one measurement instead of plotting all five measurements, because plotting five measurements would over-represent the study in the graph. Because of the low number of studies included in our systematic review and the considerable methodological as well as clinical differences between the studies, we did not calculate summary estimates of sensitivity and specificity and thus could not construct a summary ROC curve. From the ROC space in Figure 2 appears that for the individual studies for a sensitivity of 100%, specificity was around 40% whereas at a sensitivity of 60%, the specificity was at most 80%. When specificity approached 100%, sensitivity dropped to 20 and 30%.
Comment

In this systematic review we evaluated accuracy measures of the RGT to assess its suitability for diabetes screening in pregnancy. The sample of available studies was small and showed considerable heterogeneity. The studies differed from each other on several pertinent aspects, such as study design, inclusion criteria and threshold values for an abnormal RGT result (Table 1). Therefore we did not calculate summary estimates of accuracy measures or construct a summary ROC curve. Due to different timing of screening as well as different threshold values and patient selection it is impossible to directly compare the studies. We feel that based on the individual studies included in our systematic review, the sensitivity and specificity of the RGT are insufficient to use the test as a screening test for GDM. These results are in line with recommendations from international guidelines. This systematic review does not provide evidence on the potential benefit of screening to reduce perinatal and maternal complications of GDM.

We performed an extensive literature search in various databases without language or any other restrictions. We assume that we identified all articles that report on the RGT, although studies that did not mention the RGT in title, abstract or key words might have been missed in our electronic search. Three relevant studies that were not identified by our electronic search appeared to be not included in Medline or Embase. We were unable to obtain the full text of one of these studies. Our attempt to contact the author of the manuscript was not successful. However, since the results of the studies that were included in this review are already heterogeneous, we feel that results of the untraceable manuscript would probably not have changed our conclusion. An

Table 1. Key characteristics of the included studies

<table>
<thead>
<tr>
<th>First author, Year</th>
<th>Country</th>
<th>N</th>
<th>Inclusion criteria</th>
<th>Gestational age (wks)</th>
<th>Verification OGTT (%)</th>
<th>GDM n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jowett 1987</td>
<td>UK</td>
<td>110</td>
<td>Risk factors</td>
<td>27 - 31</td>
<td>100</td>
<td>49 (45)</td>
</tr>
<tr>
<td>Nasrat 1988</td>
<td>Kuwait</td>
<td>276  (250)</td>
<td>All women</td>
<td>28 - 32</td>
<td>91</td>
<td>50 (20)</td>
</tr>
<tr>
<td>Mathai 1994</td>
<td>India</td>
<td>232</td>
<td>All women</td>
<td>26 - 30</td>
<td>100</td>
<td>11 (4.7)</td>
</tr>
<tr>
<td>Tam 2000</td>
<td>China</td>
<td>895</td>
<td>All women</td>
<td>24 - 32</td>
<td>100</td>
<td>108 (12)</td>
</tr>
<tr>
<td>Maegawa 2003</td>
<td>Japan</td>
<td>749  (735 b)</td>
<td>All women</td>
<td>1st &amp; 2nd trimester</td>
<td>100</td>
<td>1.9/1.1 b (14/8 b)</td>
</tr>
<tr>
<td>Van Leeuwen 2007</td>
<td>The Netherlands</td>
<td>1301</td>
<td>All women</td>
<td>24 - 28</td>
<td>100 c</td>
<td>48 (3.8)</td>
</tr>
</tbody>
</table>

a. Study in which RGT was performed more than once; b. Number of women analyzed, and % GDM in 2nd trimester; c. After correction for verification bias, description in Appendix A; d. Fasting glucose value, glucose at 1 and 2 hours after glucose load (GDM if 2 or more conditions are met: (1) fasting glucose 5.6 mmol/L; (2) glucose 1 h after 75-g OGTT 10 mmol/L; (3) glucose 2 h after 75-g OGTT 8.3 mmol/L);
explanation for the limited number of studies that we found with our literature search could be publication bias. Publication bias occurs when studies with positive results are more likely to be submitted, or accepted for publication, than studies with negative or inconclusive results.\textsuperscript{26} If publication bias is present, the accuracy of the RGT reported in this systematic review was most likely overestimated.

Figure 2. ROC plot of the included studies for various thresholds the RGT. Displayed are the studies of Mathai (Δ), Nasrat (■), Tam (○), van Leeuwen et al. (■), Maegawa et al. (○), first trimester and O, second trimester), Jowett (▲).

<table>
<thead>
<tr>
<th>First author, Year</th>
<th>Country</th>
<th>N</th>
<th>Inclusion criteria</th>
<th>Gestational age (wks)</th>
<th>Verification OGTT (%</th>
<th>GDM n (%)</th>
<th>OGTT (gram)</th>
<th>Cut off OGTT (mmol/L</th>
<th>Time RGT - OGTT</th>
<th>Time to last meal</th>
<th>Blood glucose</th>
<th>Cut off RGT (mmol/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jowett 22, 1987</td>
<td>UK</td>
<td>110</td>
<td>Risk factors</td>
<td>27 - 31</td>
<td>100</td>
<td>49</td>
<td>75</td>
<td>8.0</td>
<td>Same day</td>
<td>Fixed times</td>
<td>Venous plasma</td>
<td>5.6 and 6.1</td>
</tr>
<tr>
<td>Nasrat 20, 1988</td>
<td>Kuwait</td>
<td>276</td>
<td>All women</td>
<td>28 - 32</td>
<td>91</td>
<td>50</td>
<td>75</td>
<td>8.0</td>
<td>&lt; 5 days</td>
<td>Independent</td>
<td>Venous plasma</td>
<td>5.8 and 6.9</td>
</tr>
<tr>
<td>Mathai 13, 1994</td>
<td>India</td>
<td>232</td>
<td>All women</td>
<td>26 - 30</td>
<td>100</td>
<td>11</td>
<td>100</td>
<td>5.3; 10; 8.6; and 7.8</td>
<td>&lt; 2 weeks</td>
<td>Independent</td>
<td>Venous plasma</td>
<td>4.4 to 6.4</td>
</tr>
<tr>
<td>Tam 21, 2000</td>
<td>China</td>
<td>895</td>
<td>All women</td>
<td>24 - 32</td>
<td>108</td>
<td>12</td>
<td>75</td>
<td>7.8</td>
<td>&lt; 4 weeks</td>
<td>Independent</td>
<td>Venous plasma</td>
<td>4.7</td>
</tr>
<tr>
<td>Maegawa 12, 2003</td>
<td>Japan</td>
<td>749</td>
<td>All women</td>
<td>1st &amp; 2nd trimester</td>
<td>100</td>
<td>1.9/1.1</td>
<td>75</td>
<td>5.6; 10; and 8.3</td>
<td>2 to 4 weeks</td>
<td>Independent</td>
<td>Venous plasma</td>
<td>5.3 and 5.6</td>
</tr>
</tbody>
</table>

\textsuperscript{a}. Fasting glucose value, glucose value at 1, 2, and 3 hours after the glucose load (GDM was diagnosed if 2 values exceeded threshold values); \textsuperscript{b}. Peak value of glucose measurements at 1, 2, and 3 hours after the glucose load (GDM was diagnosed if the threshold value of 8 mmol/L was exceeded); \textsuperscript{c}. Two hour glucose value.
The clinical applicability of a test depends, amongst others, on the probability of disease that needs further testing (or treatment). The extent to which the probability of GDM is increased or decreased compared to the probability prior to testing (pre-test probability) depends on the accuracy of the screening test as well as on prevalence of GDM in the population in which the screening test is being used. The pre-test probability depends amongst others on characteristics of the population. An approach in which the RGT could be clinically useful despite moderate measures of sensitivity and specificity is if the probability of GDM after testing exceeds the probability of disease that is required to warrant further testing (or treatment). Suppose that prevalence of GDM in a population is 3%. Assuming that the RGT has a sensitivity of 60% with a specificity of 80%, a positive result of the RGT changes the probability of GDM from 3 to 8.5%, whereas a negative result of the RGT would change the probability of GDM from 3 to 1.5%. These probabilities both are low. In a population with a prevalence of GDM of 15% however (e.g. women with risk factors for GDM), a positive result of the RGT changes
the probability of disease from 15 to 35%, whereas a negative RGT results changes the probability of disease from 15 to 8.1%. If the threshold to perform further testing is set for example at a probability of 20%, the RGT could be clinically useful in the second population, but not in the first population.

The RGT is a relatively easy and fast procedure to screen for GDM. It requires no specific preparation as for example fasting or ingestion of an oral glucose load and the test itself carries little inconvenience for women. The relative convenience and low costs of the test might be the reason for its frequent use, albeit not supported by international guidelines. If a test is used in a screening setting however, in general high sensitivity of the test is mandatory, irrespective of other characteristics such as convenience or costs. The study by Maegawa was the only study that found a more favorable combination of sensitivity and specificity for the RGT in the first trimester. Thresholds for an abnormal RGT were set at 5.3 mmol/L and 5.6 mmol/L, which could explain the relative high sensitivity, since 5.3 mmol/L is often used as a cut off value for the fasting glucose test. Furthermore, women with an abnormal OGTT in the first trimester are more likely to have diabetes type one or two that is discovered in pregnancy, rather than GDM. To evaluate the accuracy of the test to screen for GDM, accuracy of the test in second trimester should be considered. Sensitivity of the test in the second trimester was considerably lower than in the first trimester.

The study by Jowett et al. showed that the performance of the RGT is associated with timing of the test. The sensitivity of the RGT in their study ranged from 25 to 47% for random blood glucose measurement in the same women at different times of day. As pregnancy progresses plasma glucose levels under fasting conditions drop whereas plasma glucose levels after a meal become higher. As the RGT is performed at a random point in time, peak values after a meal might remain undetected. Indeed women may have normal blood glucose values with random glucose testing, but still have unnoticed (asymptomatic) periods of hyperglycemia. These peaks might contribute to adverse outcomes in pregnancy and complications during delivery. Combs et al. showed that fetal macrosomia was related to increased postprandial glucose levels. A series of RGT measurements could partly resolve this issue of variation in blood glucose values. If random glucose measurement is, for example, performed five times a day, using the highest blood glucose value as the result of the RGT, sensitivity of the RGT might be improved, though the relatively easy and convenient character of the test would be lost. A large cohort study by Ostlund and Hanson evaluated accuracy measures of multiple RGT measurements throughout pregnancy. The highest value of the RGT measured during pregnancy cross classified against the result of the OGTT resulted in a maximum sensitivity of 75.4% with a corresponding specificity of 77.9% (threshold value 6.5
mmol/L). By measuring random blood glucose values on a regular basis in pregnancy the discriminative capacity of the RGT might thus be increased.

In conclusion, based on the studies included in our systematic review, sensitivity and specificity of the RGT seem to be not sufficient to be used as a screening test. Therefore, we consider a single RGT measurement an inadequate method to screen for GDM. The potential value of the RGT in screening strategies in which individual pre-test probabilities based on, for example, patient characteristics are combined with test accuracy measures could be evaluated in decision analysis models. If, however, performance of the RGT then is not increased, the RGT has little value for detecting GDM.

Acknowledgements

We thank Dr. M. Berg (Sweden), Dr. M. Hatem (UK), Dr. W.H. Tam (China), Dr. E. Stenhouse (UK) and Dr. M. Janghorbani (UK), Professor Dr. V. Seshiah (India) and Dr. Sugiyama (Japan) for their effort to provide extra information on their studies. We thank Professor Taixiang Wu (China) for his help in extracting data from Chinese articles.
References


Appendix A. Verification bias

When a screening test is evaluated against a reference test, ideally all participating patients should undergo both tests. However, in practice, this is seldom done due to high costs of the procedure or burden to the patient. If only patients with verified screening test results are used to assess the performance of the screening test, calculated accuracy measures become biased since patients with verified disease status are only patients with abnormal screening test results, which is not a representative random sample of the population in which the screening test is being used. The bias that is introduced is called (partial) verification bias.

In the study by van Leeuwen et al., the reference test was not performed in all patients (predefined protocol). The following procedure was used 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 results of the OGTT measurements in women who were not subjected to an OGTT were estimated based on results of the random test and the 50-g glucose screening test as well as on patient characteristics using multiple logistic regression analysis. This procedure is called imputation.