Screening for gestational diabetes mellitus
van Leeuwen, M.

Citation for published version (APA):
van Leeuwen, M. (2012). Screening for gestational diabetes mellitus
Fasting glucose measurement to detect gestational diabetes mellitus
A systematic review

Marsha van Leeuwen
Helene van Meir
Mireille J. Serlie
Brent C. Opmeer
Ben Willem J. Mol

Submitted
Abstract

BACKGROUND Treatment of gestational diabetes mellitus (GDM) reduces the risk of perinatal and maternal complications. It is unclear what is the best strategy to identify women with GDM.

OBJECTIVE To review the literature on fasting glucose measurement to detect GDM.

METHODS We systematically searched Medline, Embase and reference lists of included articles to identify studies that compared fasting glucose measurement to the reference standard to diagnose GDM (either 75-g or 100-g oral glucose tolerance test (OGTT)) before 32 weeks of gestation. We scored study characteristics and quality, and extracted data to construct two-by-two tables cross-classifying results of fasting glucose measurement and the OGTT. Meta-analysis was performed using a bivariate model to estimate summary estimates of sensitivity and specificity and 95% confidence intervals for ranges of threshold values of fasting glucose measurement accounting for various test characteristics. Summary receiver operating characteristics (sROC) curves were generated.

RESULTS We included 16 studies (25 560 women, 12.1% GDM). There was no association between study population (consecutive or selective recruitment), threshold of OGTT (high or low) and summary estimates of sensitivity and specificity of fasting glucose measurement. Summary estimates of sensitivity varied from 0.30 (95% CI 0.09 - 0.65) to 0.92 (95% CI 0.81 - 0.97) (threshold value > 5.0 mmol/l and < 4.6 mmol/l respectively). Summary estimates of specificity varied from 0.96 (95% CI 0.90 - 0.98) to 0.45 (95% CI 0.27 - 0.65) for these threshold ranges.

CONCLUSION Accuracy of fasting glucose measurement is not sufficient to replace the OGTT in the diagnostic work-up for GDM.
Introduction

Gestational diabetes mellitus (GDM) is defined as glucose intolerance of varying severity with onset or first recognition during pregnancy. GDM complicates 2 to 9% of all pregnancies and the prevalence is rising due to the increasing epidemic of overweight and obesity. Women with GDM are at higher risk of pregnancy complications and adverse maternal and fetal outcomes. In addition, women with GDM have up to 60% risk of developing diabetes mellitus (type 2) within 5 to 15 years after delivery. Treatment of women with GDM improves perinatal as well as maternal outcome. Detection of GDM however is complicated, since it often is an asymptomatic condition. A screening program might facilitate detection of GDM and therefore contribute to the reduction of perinatal and maternal complications. Currently a variety of tests and methods are used in the screening and diagnostic work-up of GDM. There is no consensus on the optimal strategy. One of the tests that are used is fasting glucose measurement.

Fasting glucose measurement is inexpensive and relatively easy to perform. However, accuracy measures of the test reported in the literature are divergent. In order to assess the suitability of fasting glucose measurement in diagnostic work up of GDM we need to be well informed on test characteristics such as test accuracy. We conducted a systematic review and performed meta-analysis using bivariate regression analysis to obtain summary estimates of sensitivity and specificity in order to assess the diagnostic accuracy of fasting glucose measurement for detection of GDM.

Material and Methods

Literature search
A systematic search of the electronic databases MEDLINE (1950 to October 2011) and EMBASE (1980 to October 2011) was performed. We searched MEDLINE and EMBASE broadly for the target disease (GDM), target test (fasting glucose measurement) and population (pregnant women) using both free-text words and Subject Headings specific to each database. No methodological filter was applied as this can lead to omission of relevant papers. No language restrictions were applied. We refined the search strategy through adding search terms as new relevant citations were identified. We checked reference lists of relevant studies to identify cited articles not captured by electronic searches. Reference Manager 11.0 was used to establish a database incorporating results of all searches (Thomson ISI Research Soft, Carlsbad, CA, USA).

Study selection
Titles and abstracts of the articles identified with the systematic search were screened by two independent reviewers (M.v.L. and H.v.M.). The reviewers were not blinded to
details of the studies (e.g. author and journal). Studies were selected if they reported on fasting glucose measurement in pregnant women according to at least one of the reviewers. Full text reports were read, and studies were selected if they met the following criteria: Comparison of fasting glucose measurement (index test) to the 75 or 100-g OGTT (reference test) in pregnant women before the 32nd week of gestation at any level of risk for GDM. Data should be available to construct a 2x2 table of test performance. In cases where this was not possible, but where original data would allow generation of a two-by-two table, authors were contacted by e-mail and/or mail. We excluded diagnostic case control studies in which women with GDM were compared to women without GDM, as estimates of test accuracy are often overoptimistic in these studies.12,13 Final in-/exclusion decisions were made by comparison of the results of both reviewers. Disagreements were resolved by consensus or arbitration by a third reviewer (B.W.M.).

Data extraction
The two reviewers extracted data from English language manuscripts, using a piloted data extraction form. For other language manuscripts data was extracted by (medical) colleagues with knowledge of the particular language, using the same data extraction forms. Descriptive data (first author, country and year of publication), data on study characteristics, study quality and test accuracy results were extracted, including sample characteristics, test characteristics and study design. Data on test accuracy was abstracted as 2x2 tables cross classifying the results of fasting glucose measurement with the results of the OGTT for various threshold values of fasting glucose measurement. In case of multiple publications, all publications were used to acquire complete data. The most recent and complete results were included in the analysis. If data were missing on test accuracy or on other relevant characteristics, we contacted the corresponding author by e-mail or by mail to seek their assistance in data extraction or to obtain the missing data. We used the QUADAS tool to assess the methodological quality of the selected studies. QUADAS is a tool for the quality assessment of studies of diagnostic accuracy.14 Included studies were evaluated on 15 items concerning design, verification, patient selection, description of the tests and of the study population. Disagreement was resolved by consensus. If no consensus was reached, a third independent reviewer (B.W.M.) was consulted.

Diagnosis of GDM
GDM is diagnosed with an OGTT (reference standard). Currently the 75-g as well as the 100-g OGTT are used as diagnostic test for GDM. We therefore included studies that used either the 75-g OGTT or the 100-g OGTT as reference test. In the past, results of the OGTT were classified as normoglycemic, impaired glucose tolerance (IGT), or as GDM. Currently, the category IGT is not used anymore. To facilitate comparison between the studies and enabling meta-analysis, we considered women with IGT as either being normoglycemic or as having GDM, according to currently used criteria if necessary.
Data synthesis
Per study we calculated sensitivity and specificity with 95% confidence intervals for all thresholds of fasting glucose measurement for which data were available. We created forest plots and plotted their results in a receiver-operating characteristics (ROC) plot (sensitivity vs. 1-specificity) to explore heterogeneity.

Next, we obtained summary estimates of sensitivity and specificity and their 95% confidence intervals with bivariate regression analysis for ranges of threshold values of fasting glucose measurement.\textsuperscript{15} With a bivariate regression model summary estimates can be calculated simultaneously for sensitivity and specificity with a single model, while accounting for potential (often negative) correlations between sensitivity and specificity. Sensitivity and specificity are often negatively correlated across studies due to implicit variation of threshold values. Variation or heterogeneity in results between studies can be the result of chance variation, difference in threshold values or different clinical subgroups. The bivariate regression model also allows random effects in order to accommodate clinical or statistical heterogeneity. The bivariate regression approach can be extended with covariates to examine whether they have an effect on sensitivity, specificity or both. We performed the following predefined subgroup analyses: type of reference test (75 or 100-g OGTT), recruitment of women (consecutive / risk factors). As multiple criteria for an abnormal OGTT exist for the 75-g OGTT as well as for the 100-g OGTT, due to which comparison of studies is complicated, we categorised threshold values of the 75 and 100-g OGTT to define GDM, as being “high” or “low” in order to facilitate comparison. This classification was added as a covariate to the bivariate regression model (high threshold OGTT / low threshold OGTT).

The included studies reported different values of fasting glucose measurement to define a test result as abnormal, and several studies reported on multiple threshold values. We did not limit our analyses to one single threshold value in order to evaluate pooled estimates of sensitivity and specificity of fasting glucose measurement over the whole range of possible thresholds, by assuming that the shift in accuracy (negative correlation between sensitivity and specificity) was accounted for by the correlation term as specified in the bivariate regression model. To avoid biased results towards studies that reported on multiple threshold values, we estimated the model in 250 stratified bootstrap samples, in which each time only one threshold value from each study was randomly selected. All parameter estimates were averaged over 250 model results to calculate pooled estimates for sensitivity and specificity. Based on these estimates summary ROC curves were obtained.\textsuperscript{16} For the overall analysis, the original data concern accuracy for different threshold values to define an abnormal fasting glucose measurement, and this sROC curves reflects variation in sensitivity and specificity associated with the shift in threshold values. For the analysis per threshold, the sROC point estimate reflects the
pooled estimate for sensitivity and specificity for the respective threshold, and parameter uncertainty is indicated with a 95% confidence region. Statistical analyses were performed using SPSS 16.0 (SPSS Inc., Chicago, IL, USA) and SAS 9.1.3. (Proc NLMixed). Forest plots were made with Review Manager (RevMan) [Computer program]. Version 5.0. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2008.

Results

Figure 1 shows the results of our search and selection of studies. Our search resulted in 1005 hits. 222 studies were selected for further reading based on title and abstract. After assessment of full text reports we excluded 206 reports for various reasons. The remaining 16 reports were included in the systematic review (Table 1). The studies reported on a total of 25,560 women of whom 3,096 were diagnosed with GDM (12.1%).

Study characteristics

Characteristics of included studies are summarised in Table 1. Fourteen studies were cohort studies (88%), 2 studies had a cross sectional design (12%). Data collection was prospective in 15 studies (94%). Fourteen studies reported consecutive recruitment of patients (88%). Sample sizes ranged from 123 women to 4,528 women (median 887 women). The incidence of GDM ranged from 1.1 to 31.0% (median 16.2%).

Reference test

In three studies (1,672 women) the 100-g OGTT was used as a reference test for GDM, whereas in 13 studies (23,888 women) the 75-g OGTT was used as a reference test. Ten studies were categorised as having low OGTT threshold, and four studies were categorised as having high OGTT threshold. In two studies there were multiple criteria for GDM (both high and low).

Index test

The number of thresholds of fasting glucose measurement in the individual studies varied between one and 11. Thresholds ranged from 3.9 mmol/l to 7.0 mmol/l. There was no threshold of fasting glucose measurement for which all studies reported accuracy measures. The threshold that was most frequently applied was 5.0 mmol/l. In the majority of the studies fasting glucose measurement and the OGTT were performed between 24 and 28 weeks of gestation. In two studies the tests were performed throughout the whole period of pregnancy, but mostly before 34 weeks of gestation. The period of fasting was most often described as “overnight fasting”.

Quality assessment

Selection criteria were reported in all studies (100%). In 14 studies (88%) complete verification of the index test was performed. In two studies (12%) more than 75% of
Possible relevant studies retrieved from the electronic databases
N = 1005

Citations excluded after screening titles and abstracts
N = 783

Relevant studies retrieved for evaluation of the full text
N = 222

Eligible articles from reference lists / the Web of Science
N = 0

Relevant studies retrieved for evaluation of the full text
N = 222

Excluded studies
- No data on the FGT and OGTT N = 82
- Other screening test than FGT performed N = 50
- Editorial / review report; no data N = 9
- No data reported in article N = 30
- Data already reported in other article N = 4
- Not only pregnant women included N = 1
- Only women with abnormal FGT and normal OGTT N = 1
- Only women with GDM included N = 7
- Differential verification (only when FGT positive) N = 6
- Other bias N = 3
- No full text available N = 3
- Case control design N = 9
- Other reference test than OGTT N = 1
- Total excluded N = 206

Studies included in the systematic review
N = 16

Figure 1. Results of literature search and selection process.

the index tests were verified. Details on description of the index and reference test were reported in 13 studies (81%). In 3 studies (19%) fasting glucose measurement and OGTT were performed separately. In 12 studies (75%) fasting glucose measurement was part of the OGTT. In one study (6%) it was unclear if fasting glucose measurement and OGTT were performed separately.

Data analysis
Figure 2 shows the combination of sensitivity and specificity of all threshold values from all included studies. Sensitivity of individual studies varied widely from 3 to 100% and specificity from 0 to 100%, mainly due to variation in threshold values of fasting glucose measurement (range 3.9 mmol/l to 7.0 mmol/l). With the bivariate regression model we calculated pooled summary estimates of sensitivity and specificity for ranges of threshold values of fasting glucose measurement. The bootstrapped overall sROC curve
is displayed in Figure 2. Similar analyses on subgroups defined by type of reference test (75 / 100-g OGTT), recruitment (consecutive / selective), threshold of the reference test (high / low) did not suggest differences in accuracy across these groups (results not shown). Figure 3 shows the summary point estimates for sensitivity and specificity with 95% CIs for various ranges of threshold values of fasting glucose measurement (>5.0 mmol/l, 4.6-5.0 mmol/l and <4.6 mmol/l). For these respective threshold ranges, pooled estimates of sensitivity varied from 0.30 (95% CI 0.09 - 0.65) to 0.92 (95% CI 0.81 - 0.97), and for specificity from 0.96 (95% CI 0.90 - 0.98) to 0.45 (95% CI 0.27 - 0.65) (Table 2).

---

### Table 1. Characteristics of the included studies

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Country</th>
<th>Design</th>
<th>Inclusion, recruitment</th>
<th>OGGT (gram)</th>
<th>Threshold OGGT</th>
<th>FGT and OGGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agarwal, 2006</td>
<td>UAE</td>
<td>Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>High / low</td>
<td>One test</td>
</tr>
<tr>
<td>Agarwal, 2005</td>
<td>UAE</td>
<td>Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
</tr>
<tr>
<td>Agarwal, 2000</td>
<td>UAE</td>
<td>Cohort</td>
<td>Prospective, consecutive</td>
<td>100</td>
<td>Low</td>
<td>One test</td>
</tr>
<tr>
<td>Bito, 2005</td>
<td>Hungary</td>
<td>Cohort</td>
<td>Prospective, Risk factors</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
</tr>
<tr>
<td>Fadl, 2006</td>
<td>Sweden</td>
<td>Cross sectional</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>High</td>
<td>Separate</td>
</tr>
<tr>
<td>Kauffman, 2006</td>
<td>USA</td>
<td>Cohort</td>
<td>Prospective, consecutive</td>
<td>100</td>
<td>High / low</td>
<td>One test</td>
</tr>
<tr>
<td>Maegawa, 2003</td>
<td>Japan</td>
<td>Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>High</td>
<td>One test</td>
</tr>
<tr>
<td>Perucchini, 1999</td>
<td>Switzerland</td>
<td>Cohort</td>
<td>Prospective, consecutive</td>
<td>100</td>
<td>Low</td>
<td>One test</td>
</tr>
<tr>
<td>Reichelt, 1998</td>
<td>Brazil</td>
<td>Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
</tr>
<tr>
<td>Rey, 2004</td>
<td>Canada</td>
<td>Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>High</td>
<td>Separate</td>
</tr>
<tr>
<td>Sacks, 2003</td>
<td>USA</td>
<td>Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>High</td>
<td>Unknown</td>
</tr>
<tr>
<td>Sayeed, 2005</td>
<td>Bangladesh</td>
<td>Cross sectional</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
</tr>
<tr>
<td>Senanayake, 2006</td>
<td>Sri Lanka</td>
<td>Cohort</td>
<td>Prospective, Risk factors</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
</tr>
<tr>
<td>Seshiah, 2004</td>
<td>India</td>
<td>Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
</tr>
<tr>
<td>Tam, 2000</td>
<td>China</td>
<td>Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>Low</td>
<td>Separate</td>
</tr>
<tr>
<td>Wijayarathne, 2006</td>
<td>Sri Lanka</td>
<td>Cohort</td>
<td>Retrospective, consecutive</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
</tr>
</tbody>
</table>

* Range of GA at testing unknown, mean GA at testing was 26 weeks; ** Also screening in first trimester, women with hyperglycemia in the first trimester were excluded; *** Majority in 2nd trimester
Table 1. Characteristics of the included studies

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Country</th>
<th>Design</th>
<th>Inclusion, recruitment</th>
<th>OGTT (gram)</th>
<th>Threshold OGTT</th>
<th>FGT and OGTT</th>
<th>Gestational age (weeks)</th>
<th>Venous / capillary and plasma / blood</th>
<th>Sample size</th>
<th>GDM (n) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agarwal, 2006</td>
<td>UAE Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>High / low</td>
<td>One test</td>
<td>24 - 28</td>
<td>Venous, plasma</td>
<td>4602</td>
<td>979 (21.6)</td>
<td></td>
</tr>
<tr>
<td>Agarwal, 2005</td>
<td>UAE Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
<td>24 - 28</td>
<td>Venous, plasma</td>
<td>1685</td>
<td>333 (19.8)</td>
<td></td>
</tr>
<tr>
<td>Agarwal, 2000</td>
<td>UAE Cohort</td>
<td>Prospective, Risk factors</td>
<td>100</td>
<td>Low</td>
<td>One test</td>
<td>24 - 28</td>
<td>Plasma</td>
<td>1276</td>
<td>396 (31.0)</td>
<td></td>
</tr>
<tr>
<td>Bito, 2005</td>
<td>Hungary Cohort</td>
<td>Prospective, risk factors</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
<td>24 - 28</td>
<td>Venous, plasma</td>
<td>155</td>
<td>32 (20.7)</td>
<td></td>
</tr>
<tr>
<td>Fadl, 2006</td>
<td>Sweden</td>
<td>Cross sectional</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>High Separate</td>
<td>28 - 32</td>
<td>Capillary</td>
<td>3610</td>
<td>55 (1.5)</td>
<td></td>
</tr>
<tr>
<td>Kauffman, 2006</td>
<td>USA Cohort</td>
<td>Prospective, consecutive</td>
<td>100</td>
<td>High / low</td>
<td>One test</td>
<td>24 - 28</td>
<td>Venous, plasma</td>
<td>123</td>
<td>16 (13.0) / 25 (20.3)</td>
<td></td>
</tr>
<tr>
<td>Maegawa, 2003</td>
<td>Japan Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>High</td>
<td>One test</td>
<td>24 - 28</td>
<td>** Unknown</td>
<td>735</td>
<td>8 (1.1)</td>
<td></td>
</tr>
<tr>
<td>Perucchini, 1999</td>
<td>Switzerland Cohort</td>
<td>Prospective, consecutive</td>
<td>100</td>
<td>Low</td>
<td>One test</td>
<td>24 - 29</td>
<td>Venous, plasma</td>
<td>520</td>
<td>53 (10.2)</td>
<td></td>
</tr>
<tr>
<td>Reichelt, 1998</td>
<td>Brazil Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
<td>24 - 28</td>
<td>Plasma</td>
<td>5010</td>
<td>379 (7.6)</td>
<td></td>
</tr>
<tr>
<td>Rey, 2004</td>
<td>Canada Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>High Separate</td>
<td>24 - 28</td>
<td>Venous, plasma</td>
<td>188</td>
<td>21 (11.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sacks, 2003</td>
<td>USA Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>High</td>
<td>Unknown</td>
<td>12 - 34</td>
<td>Venous, plasma</td>
<td>4507</td>
<td>302 (6.7)</td>
<td></td>
</tr>
<tr>
<td>Sayeed, 2005</td>
<td>Bangladesh</td>
<td>Cross sectional Prospective, consecutive</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
<td>24 - 28</td>
<td>Capillary, blood</td>
<td>147</td>
<td>12 (8.2)</td>
<td></td>
</tr>
<tr>
<td>Senanayake, 2006</td>
<td>Sri Lanka Cohort</td>
<td>Prospective, Risk factors</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
<td>26</td>
<td>Plasma</td>
<td>271</td>
<td>75 (27.7 *)</td>
<td></td>
</tr>
<tr>
<td>Seshiah, 2004</td>
<td>India Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
<td>2nd-3rd ** ** ** trimester</td>
<td>Venous, plasma</td>
<td>520</td>
<td>53 (10.2)</td>
<td></td>
</tr>
<tr>
<td>Tam, 2000</td>
<td>China Cohort</td>
<td>Prospective, consecutive</td>
<td>75</td>
<td>Low Separate</td>
<td>28 - 32</td>
<td>Capillary, plasma</td>
<td>1031</td>
<td>122 (11.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wijeyaratne, 2006</td>
<td>Sri Lanka Cohort</td>
<td>Retrospective, consecutive</td>
<td>75</td>
<td>Low</td>
<td>One test</td>
<td>24 - 28</td>
<td>Capillary, plasma</td>
<td>883</td>
<td>144 (16.3)</td>
<td></td>
</tr>
</tbody>
</table>

* Range of GA at testing unknown, mean GA at testing was 26 weeks; ** Also screening in first trimester, women with hyperglycemia in the first trimester were excluded; *** Majority in 2nd trimester

Table 2. Summary estimates for sensitivity and specificity of fasting glucose measurement calculated with bivariate meta-analysis.

<table>
<thead>
<tr>
<th>Threshold (mmol/l)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 4.6 mmol/l</td>
<td>0.92 (0.81 - 0.97)</td>
<td>0.45 (0.27 - 0.65)</td>
</tr>
<tr>
<td>4.6 - 5.0 mmol/l</td>
<td>0.75 (0.60 - 0.86)</td>
<td>0.70 (0.47 - 0.86)</td>
</tr>
<tr>
<td>&gt; 5.0 mmol/l</td>
<td>0.30 (0.09 - 0.65)</td>
<td>0.96 (0.90 to 0.98)</td>
</tr>
</tbody>
</table>
Figure 2. Summary Receiver Operating Characteristic (sROC) plot with summary estimate of sensitivity and specificity (*) for all threshold values of fasting glucose measurement and all possible values of the Oral Glucose Tolerance Test (OGTT) (n=86).

Figure 3. sROC plot for various thresholds of fasting glucose measurement. Cutoff 1 = threshold fasting glucose measurement < 4.6 mmol/l Cutoff 2 = threshold fasting glucose measurement 4.6 mmol/l – 5.0 mmol/l Cutoff 3 = threshold fasting glucose measurement > 5.0 mmol/l
Discussion

We performed a systematic review of the literature with meta-analysis in order to assess whether fasting glucose measurement is suitable to use in the diagnostic work up of GDM. There were 16 studies reporting on 25 560 women that met the inclusion criteria. We found that summary estimates of sensitivity and specificity were not different for the 75-g or the 100-g OGTT and the threshold (high or low) and recruitment (consecutive or inclusion of women with risk factors). The best combination of summary estimates of sensitivity and specificity was for a threshold value of fasting glucose measurement of 4.6-5.0 mmol/l (summary estimate of sensitivity 0.75 (95% CI 0.60 - 0.86) and summary estimate of specificity 0.70 (95% CI 0.47 - 0.86), which is not sufficient to replace the OGTT in the diagnostic work-up for GDM.

We used the QUADAS tool for assessment of methodological quality of individual included studies, thereby identifying studies with severe methodological limitations. Although we did not perform subgroup analyses to formally assess the influence of the study quality on the accuracy measures, we were able to assess the quality of all individual studies in order to exclude studies that did not meet the major requirements. We were limited to available primary studies and data that were reported in these studies. If individual studies selected threshold values of the index study associated with high accuracy measures (data driven selection of the optimal threshold value) this causes overoptimistic estimates of sensitivity and specificity. It was not possible to perform subgroup analysis for certain variables as for example type of blood analysis (venous or capillary and plasma or whole blood) because there were too little data. Only one study used blood glucose, where others used plasma glucose.

Bivariate meta-analysis is a valid method to estimate summary estimates of accuracy measures of a diagnostic test. However, one depends on data that is reported by the authors in the original articles. Not all potential covariates or subgroups can be accounted for, possibly leading to results that are not applicable for the general population. One way to overcome this is by performing individual patient data analysis (IPD), which allows treatment effects and diagnostic accuracy to be estimated at the level of for example specific subgroups.

Fasting glucose measurement is an appealing test for detection of GDM because of its convenience. It is well tolerated and has relative low costs. On the other hand, normal fasting glucose levels do not exclude postprandial hyperglycemia, which is thought to affect fetal overgrowth. Postprandial hyperglycemia can be excluded by a glucose-loading test. An adequate screening test should have a sufficient degree of reproducibility. There have been no studies on the reproducibility of fasting glucose measurement. Sacks
et al. found that in women with GDM as well as in women without GDM fasting glucose values rise when pregnancy progresses. The day-to-day reproducibility has to our knowledge never been studied. Moreover, the length of fasting is a determinant of fasting glucose because endogenous glucose production decreases over time.

In most studies accuracy of fasting glucose measurement as screening test for GDM is evaluated by comparing results of fasting glucose measurement to results of the OGTT (reference standard). The original criteria to define the OGTT as abnormal were set to identify women at risk for developing diabetes (type two) in the future and not to select women at risk for adverse perinatal or maternal outcome. Preferably results of fasting glucose measurement should be compared to the risk of adverse pregnancy outcome instead of to the results of the OGTT in order to assess its suitability in the work up for GDM. In 2008 results of a large study on the association of blood glucose values of the OGTT with perinatal and maternal outcomes were published. The study showed that there is a continuous association of fasting blood glucose levels in pregnancy with a number of perinatal and maternal outcomes. Considering this association it might be plausible that fasting glucose measurement can indeed be used in the diagnostic work up for GDM. However since the association between fasting glucose values and perinatal and maternal outcomes was found to be continuous, there was no obvious thresholds at which risks increased.

An adequate test that is used in the diagnostic work up of GDM should have high sensitivity, in order to miss as little women with GDM as possible. High sensitivity is amongst others related to the threshold that is applied to define the screening test result as abnormal. With lower threshold values, higher sensitivity is achieved at the price of having to perform more OGTTs (lower specificity). Ideally an adequate test should have a high sensitivity, but not at the cost of unacceptable low specificity, since low specificity exposes a large number of women to an avoidable OGTT causing inconvenience and anxiety. However, we feel that the high sensitivity of 92% that we observed for a specificity of 45% identifies potential of this test as a first line screening test for GDM, as it prevents further testing in almost 50% of the women for a very low false-negative rate.

Further evaluation might reveal if fasting glucose measurement indeed can be used for diagnosis of GDM, and perhaps if there are specific groups that benefit from the fasting glucose measurement. Until then the results of this meta-analysis show that accuracy of fasting glucose measurement is not sufficient to replace the OGTT in the diagnostic work-up for GDM.
References


31. Tam WH, Rogers MS, Yip SK, Lau TK, Leung TY. Which screening test is the best for gestational impaired glucose tolerance and gestational diabetes mellitus? Diabetes Care 2000; 23(9):1432.


