Screening for gestational diabetes mellitus
van Leeuwen, M.

Citation for published version (APA):
van Leeuwen, M. (2012). Screening for gestational diabetes mellitus
Chapter 10

Summary, general discussion and implications
Summary

Gestational diabetes mellitus (GDM) is associated with perinatal and maternal complications,1-3 and very likely has long-term consequences for mother and child, including predisposition to obesity, metabolic syndrome and diabetes later in life.4-7 In order to allocate appropriate resources to pregnancy, perinatal management as well as to postpartum monitoring and follow up, (early) detection of women with GDM is important. The growing epidemic of obesity and diabetes mellitus type 2, that will probably result in an increasing prevalence of GDM, emphasises the need for accurate detection of GDM even more.8

An important issue in detecting GDM is that symptoms or signs are often absent. Identification of women with GDM is therefore challenging. Over the years many different strategies for screening and diagnosis of GDM have been advocated.

In 1968 Wilson and Jungner set out criteria to guide the selection of conditions suitable for screening in general. At that time, they stated that “The central idea of early disease detection and treatment is essentially simple. However, the path to its successful achievement is far from simple, though sometimes it may appear deceptively easy”.9 More than 40 years later, this statement still applies to screening for GDM. For a long time the nature and the quantity of complications associated with GDM were uncertain. Furthermore it was doubtful if treatment was beneficial in reducing these complications. In keeping with these uncertainties, there was no uniform policy in screening, diagnosis and treatment of GDM and critics questioned if screening should be performed at all.

Results of three large trials and two systematic reviews that were published recently have provided more clarity on the risks associated with GDM and the beneficial effect of treatment.1-3,10-12 Since GDM predominantly is an asymptomatic condition, the only way that women with GDM can be detected early enough for an intervention to be effective is by screening. Different tests and strategies that have been used for screening over the last years include random glucose testing, the glucose challenge test, fasting glucose measurement and selection based on patient characteristics. It is unclear what the best strategy is to identify women with GDM.

The aim of this thesis was to evaluate the accuracy and costs of various screening strategies comprising glucose tests as well as risk estimation based on patient characteristics in order to obtain an adequate strategy to detect women with GDM. We compared the performance of two screening tests with data from a cohort study and we systematically reviewed the literature on various screening tests for GDM. We validated an existing risk scoring system and developed a clinical prediction model using patient characteristics.
to estimate women’s individual risk of GDM. Furthermore we reviewed the literature on perinatal and maternal risks associated with GDM and the effect of treatment, and performed a survey amongst Dutch gynaecologists and midwives to evaluate clinical practice regarding detection and treatment of GDM in the Netherlands. Finally, we performed a cost-effectiveness analysis comparing the various screening tests and strategies.

Chapter 1 gives an outline and describes the objective of this thesis.

Chapter 2 presents the results of a study on the comparison of two screening tests for GDM: random glucose testing and the 50-g glucose challenge test. In a prospective cohort study 1301 pregnant women underwent both screening tests between 24 and 28 weeks of gestation. The reference test to diagnose GDM was a 75-g oral glucose tolerance test (OGTT). The area under the Receiver Operating Characteristics (ROC) curve of the 50-g glucose challenge test was 0.88 (95% CI 0.83 - 0.93). The area under the ROC curve of random glucose testing was 0.69 (95% CI 0.61 - 0.78). The difference between the areas under the curve was 0.19 (95% CI 0.11 - 0.27). There have been two other studies that directly compared the two screening tests in the same sample of women. McElduff et al. found their results in favour of the 50-g challenge test, whereas Mathai et al. found that the area under the ROC curve was larger for random glucose testing compared to the 50 g glucose challenge test if both tests were performed between 26 and 30 weeks of gestation. In both studies the reference test for diagnosis of GDM was a 100-g OGTT. Based on the findings of our study we conclude that the 50-g glucose challenge test is more useful than the random glucose test.

In chapter 3, 4 and 5 we systematically reviewed the literature on various screening tests for GDM and performed bivariate meta-analyses to calculate summary estimates of sensitivity and specificity when possible.

Chapter 3 reports on the results of a systematic review on the accuracy of random glucose testing as screening test for GDM. We included six studies, reporting on 3537 women. Due to small number of studies and clinical 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 a sensitivity of 60% specificity was at most 80%. When specificity approached 100%, sensitivity dropped to 20-30%. Although based on few studies with considerable clinical heterogeneity, we consider single random glucose measurement inadequate to screen for GDM.

Chapter 4 presents the results of a systematic review and bivariate meta-analysis on the accuracy of the 50-g glucose challenge test as screening test for GDM. We included
26 studies (comprising 13,564 women) that compared the 50-g glucose challenge test with the 75- or 100-g OGTT before 32 weeks of gestation. In studies that included women with risk factors for GDM pooled estimates were 0.74 (95% CI 0.62 - 0.87) for sensitivity and 0.77 (95% CI 0.66 - 0.89) for specificity (threshold value of 7.8 mmol/l). Corresponding likelihood ratios (LR) of positive and negative tests were 3.2 (95% CI 2.0 - 5.2) and 0.34 (95% CI 0.22 - 0.53) respectively. In studies with consecutive recruitment, pooled estimates were 0.74 (95% CI 0.62 - 0.87) for sensitivity and 0.85 (95% CI 0.80 - 0.91) for specificity. Derived LRs for positive and negative tests were 4.9 (95% CI 3.5 - 7.0) and 0.31 (95% CI 0.20 - 0.47). Although higher detection rates would be preferable, a detection rate of 74% seems acceptable. To use the 50-g glucose challenge test as a definite diagnostic test for GDM (replacement of the OGTT), higher accuracy measures are necessary. We conclude that the 50-g glucose challenge test is acceptable as a screening test for GDM, but cannot replace the OGTT.

In chapter 5 we systematically reviewed the literature on fasting glucose measurement as screening test for GDM. We included 16 studies that compared fasting glucose measurement to the reference standard to diagnose GDM (either 75-g or 100-g OGTT) before 32 weeks of gestation, reporting on 25,560 women. 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 calculated with a bivariate regression model were 0.30 (95% CI 0.09 - 0.65), 0.75 (95% CI 0.60 - 0.86) and 0.92 (95% CI 0.81 - 0.97) for threshold values of fasting glucose measurement of > 5.0 mmol/L, 4.6 - 5.0 mmol/L and < 4.6 mmol/L respectively. Summary estimates of specificity were 0.96 (95% CI 0.90 - 0.98), 0.70 (95% CI 0.47 - 0.86) and 0.45 (95% CI 0.27 - 0.65) for these threshold ranges. An adequate screening test should have a high sensitivity, but not at the cost of undesirable low specificity, since low specificity exposes a large number of women to an avoidable OGTT causing inconvenience and anxiety. We conclude that accuracy of fasting glucose measurement appears to be insufficient to replace the OGTT in the diagnostic work-up for GDM. Future research should reveal if fasting glucose measurement is useful for screening in specific subgroups. Possibilities of combining the 50-g glucose challenge test and fasting glucose measurement with other screening strategies should be explored.

In chapter 6 the results of a validation study are described. With data from a prospective cohort study we validated a clinical scoring system that was developed in Canada to estimate the risk of GDM. Women were assigned a score based on age, BMI and ethnicity. Performance of the scoring system was evaluated in terms of discrimination and calibration (agreement between clinical score and observed probability of GDM). The scoring system discriminated moderately (area under the curve = 0.64 (95% CI 0.56 -
0.72)) and calibration was limited. The screening strategy derived from the scoring system reduced the number of women to be tested with an OGTT with 25% for a comparable detection rate to universal screening. We conclude that despite moderate discriminative capacity and calibration of the scoring system, the screening strategy based on the scoring system appears clinically useful. However, we felt that better prediction models for GDM are needed.

Chapter 7 describes the development of a clinical prediction model for GDM that we constructed with data from a prospective cohort study. The predictive capacity as well as the clinical impact of the model was evaluated. The probability of GDM could be predicted from ethnicity, family history, history of GDM and body mass index. The prediction model had an area under the ROC curve of 0.77 (95% CI 0.69 - 0.85). Calibration was good (Hosmer and Lemeshow test statistic, p = 0.25). If an OGTT was performed in all women with a predicted probability of 2% or more, 43% of all women would be tested and 75% women with GDM would be identified. We conclude that the clinical prediction model could be useful to identify women at increased risk for GDM and that the model has the potential to improve efficiency of screening for GDM. However, the model needs to be externally validated before it can be used in clinical practice.

Chapter 8 reviews the literature on the complications associated with hyperglycemia in pregnancy and the effect of treatment of GDM. The results of three large trials are described. The studies showed that there is a linear association between glucose levels after an OGTT and the risk of maternal and perinatal complications, and that treatment of GDM reduces the risk of complications. This chapter also presents results of a survey that was performed amongst gynaecologists and midwives to describe clinical practice on screening, diagnostics and treatment of GDM in the Netherlands. The survey was performed before publication of the Dutch guideline “Diabetes and Pregnancy” in 2010. At that time, the majority of gynaecologists and midwives reported to perform screening for GDM in the first and second trimester of pregnancy. There was a large variety in tests and strategies that were used for screening. This was in line with data from surveys from other countries. The test most frequently used to perform screening was random glucose testing. The tests most frequently reported to diagnose GDM was a “breakfast” or “lunch” test (43%) followed by the 75-g OGTT (31%). We conclude that before publication of the Dutch guideline “Diabetes and pregnancy” suboptimal tests were used to detect women with GDM. We do not have data on the application of tests after publication of the guideline.

Chapter 9 presents the results of a model based cost-effectiveness analysis to evaluate which screening strategy is most cost-effective in reducing the risk of serious
complications related to GDM. A screening strategy based on a prediction model using patient characteristics combined with fasting glucose measurement and a full OGTT in case of an abnormal result of the fasting glucose measurement was the strategy associated with lowest costs to prevent serious perinatal complications (composite outcome of perinatal death, shoulder dystocia and birth trauma) (€26 172 per prevented serious perinatal complication). More complications can be prevented using more costly test strategies as universal screening with an OGTT (€43 171 per prevented serious perinatal complication) depending on the willingness to pay per prevented complication. If shoulder dystocia was excluded from the composite outcome, the strategy in which the prediction model was combined with fasting glucose measurement was still associated with lowest cost per prevented complication. However, costs per prevented outcome were higher (€65 430 per prevented complication).

**General discussion**

The work presented in this thesis focuses on screening strategies for GDM. We are aware that the debate on GDM in general, that has been going on for decades and that is foreseen not to be finished in the very near future, is more widespread than just the part on screening, and that many challenges lie ahead. Nevertheless we hope that the work described in this thesis on the evaluation of screening strategies will in the end contribute to a uniform and adequate strategy to improve pregnancy outcomes in women with GDM.

The original criteria for GDM were established by O’Sullivan and Mahan in 1964 and were based on the 100-g OGTT. These criteria initially selected women at risk for developing diabetes mellitus (type 2) in the future and did not reflect the risk for complications during pregnancy and delivery. Over the years the American Diabetes Association (ADA) and the World Health Organisation (WHO) adopted respectively the 100-g OGTT and the 75-g OGTT, each with their own specific threshold values. More recently (in 2010) the International Association of the Diabetes in Pregnancy Study Group (IADPSG) recommended new diagnostic criteria for GDM based on the 75-g OGTT. As we started our research in 2006, the work presented in this thesis is mainly based on the WHO criteria for GDM and to a lesser extent on the ADA criteria. To consider the result of our work in the context of the recently proposed IADPSG criteria as well, the background and consequences of application of these criteria will be discussed.

The IADPSG criteria were recommended after extensive analyses of the Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) trial. In the HAPO trial over 23 000 (unselected) women were subjected to a 75-g OGTT between 24 and 28 weeks of
gestation in order to define and to quantify the association between maternal glucose values and adverse pregnancy outcomes. A continuous positive association was found between maternal glucose levels (also below diagnostic criteria for GDM) and the risk of adverse pregnancy outcomes (i.e. macrosomia, primary caesarean section, clinical neonatal hypoglycemia, cord-blood serum C-peptide level > 90th percentile, pre-eclampsia, intensive neonatal care and hyperbilirubinemia). Although the results of the study were sound and clear, clinical implications were more difficult to determine because there were no clear thresholds for risk, and the outcome consisted of multiple factors related to perinatal as well as maternal outcome. After thorough international consultation, the IADPSG has set up two consensus agreements for GDM screening. The first one is straightforward: Women with risk factors should be screened early in pregnancy for unrecognised type 2 diabetes because they are at increased risk of having a fetus with congenital anomalies. The second IADPSG agreement brought along more controversy: All women should undergo screening for GDM with a 75-g OGTT at 24-28 weeks of gestation. Criteria for diagnosis of GDM are met if one of three blood glucose values exceeds specified levels (fasting >5.1 mmol/L; 1 hour >10.0 mmol/L; 2 hour >8.5 mmol/L). These threshold values are based on an odds ratio (OR) of 1.75 for the primary outcomes of the HAPO study (birth weight > 90th percentile, primary caesarean section, clinical neonatal hypoglycemia, and cord-blood serum C-peptide level > 90th percentile).

Although the IADPSG consensus agreement has been an important step towards international agreement on diagnosis of GDM, it has also caused a lot of rumour. The main concern is that if all women are subjected to an OGTT, and only one abnormal result (out of three measurements) is required for diagnosis, the number of women diagnosed with GDM will rise extensively to 17.8% compared with the WHO and ADA criteria. This will bring along substantial costs associated with screening and treatment. Furthermore, the effect of treatment of GDM based on IADPSG criteria is unknown. Although a recent meta-analysis has shown that treatment of GDM is effective in the intervention trials by Crowther and Landon, it can not be taken for granted that this effect will be confirmed in women diagnosed with GDM according to the newly proposed IADSPG criteria.

The main difference with the ADA criteria is that the new IADPSG criteria require only one instead of two abnormal results on the OGTT for diagnosis of GDM. The main difference with the WHO criteria is that with the former the 2 hour threshold value for GDM was set at 7.8 mmol/L. Although the WHO criteria and the IADPSG criteria are different, a recent meta-analysis of studies examining the WHO and IADPSG criteria demonstrated increased risk for adverse pregnancy outcomes for both criteria, with
comparable magnitudes of associations.\textsuperscript{3} For the WHO criteria, positive associations were more consistent across studies than for the IADPSG criteria.

If the IADPSG criteria are going to be adopted globally, the number of OGTTs will rise extensively. One approach to keep the number of women to be tested with an OGTT within limits could be selective screening based on risk factors and patient characteristics. This approach could be particularly useful in populations with low prevalence of GDM. If the individual risk of GDM can be predicted accurately, women at high risk for GDM should be tested with the OGTT, decreasing the burden for all women at low risk for GDM, as well as decreasing costs for society. Many studies have evaluated the use of risk factors in screening for GDM, but only few studies have summarised their results in a model or scoring system in which factors are combined and each factor is attributed its appropriate weight.\textsuperscript{15,21} We have developed a model with multiple logistic regression analysis, that can estimate the individual risk of GDM at booking based on combined patient characteristics. The risk of GDM in our cohort could be predicted from ethnicity, family history of diabetes mellitus, history of GDM and body mass index (BMI). A limitation of our study was that, in the original cohort study, the decision to perform diagnostic testing depended on the results of two screening tests generating verification bias. Although we accounted for this verification bias with a multiple imputation procedure (an accepted technique to deal with missing data), we do not know if this has influenced the result of our prediction model. The estimated detection rate of screening with our model was 75%. This is higher than in studies that assessed each maternal characteristic or risk factor as a separate screening test. The area under the receiver operating characteristics curve (AUC) was 0.77 (95\% CI 0.69 - 0.85). Nanda et al. developed a similar model to estimate the risk of GDM from patient characteristics that yielded comparable results to our prediction model (AUC of 0.79 (95\% CI 0.76 - 0.82)). They found that maternal age, BMI, ethnic origin, previous GDM or delivery of macrosomic neonates were independent predictors of GDM, with previous GDM being the strongest predictor.\textsuperscript{22} Prediction models need prospective validation before they can be used in clinical practice. The models should be tested in populations that differ from the population that was used to develop them in, to validate their use in clinical practice. If the IADPSG criteria are going to be accepted globally, validation studies should be performed preferably in a setting using these criteria. Furthermore, the extent to which screening with the prediction model may lead to improved maternal and perinatal outcome remains to be established. The importance of external validation is emphasised by our own validation study described in chapter 6. We validated a scoring system for GDM, and found that in our population discrimination and calibration of the scoring system were limited even though two out of three risk factors were included in the prediction model that we developed later on. Usually, external validation shrinks the performance of prediction models.
One of the factors that are often identified as independent risk factor for GDM is BMI. Critics believe that BMI (or rather obesity) is an independent risk factor for adverse pregnancy outcome, rather than a risk factor for GDM, and that treating obesity is more relevant than treating GDM in reducing the rate of perinatal complications. From primary and secondary analyses of the HAPO trial appears that both GDM and BMI are independent risk factors for adverse pregnancy outcome. Higher maternal BMI has been found to be associated with fetal growth and adiposity and pre-eclampsia. Since treatment strategies for obesity have rarely been successful, we feel that from clinical perspective treatment of GDM is important in order to reduce the risk of perinatal complications.

Results of screening tests as well as results of the diagnostic OGTT are often classified as “normal” or “abnormal”, depending on a pre-defined threshold value. The association between glucose levels and the risk of complications however has been proven to be continuous. No obvious threshold values for the risk of complications can be set, as the effects of maternal hyperglycemia on pregnancy outcomes do not occur at specific thresholds, but are increased on a continuum with increasing hyperglycemia. It is therefore unclear at which degree of hyperglycemia treatment should be provided. Intervention studies in women with hyperglycemia reaching the level of GDM according to WHO and ADA criteria, resulted in a reduction of perinatal complications. Results from a systematic review have shown that interventions are also effective in women with glucose concentrations below those diagnostic of GDM, by reducing the number of macrosomic babies. This clearly reflects the continuous character of GDM. Dichotomising results of glucose testing seems to be only necessary to decide which women should receive treatment. Perhaps the continuous character of risk estimation with a prediction model could reflect the continuum of risk associated with hyperglycemia. Opportunities of risk estimation should be explored by validating and improving existing prediction models possibly combined with glucose tests better tolerated than the OGTT to explore if this is an accurate method to select women who will benefit from treatment without the need to necessarily classify women as having GDM or not.

We performed systematic reviews for three individual screening tests and compared the 50-g glucose challenge test with random glucose testing. We concluded that random glucose testing is not useful in screening for GDM. Accuracy measures of the 50-g glucose challenge test and fasting glucose measurement were comparable, although somewhat in the advantage of the 50-g glucose challenge test. A model based cost effectiveness analysis showed that screening based on a prediction model using patient characteristics combined with fasting glucose measurement was the strategy associated with lowest costs to prevent serious perinatal complications (€26 172) per prevented...
serious perinatal complication), but only at the expense of a relative high rate of GDM related perinatal complications. An OGGT for all pregnant women reduced GDM related complications considerably for an acceptable cost-effectiveness ratio. So, if universal screening with an OGGT is performed, more complications can be prevented at higher cost (€43 171) per prevented serious perinatal complication) depending on the willingness to pay per prevented complication. Our cost-effective analysis was based on the WHO criteria and compared various screening strategies. Ohno et al performed a model based cost-effectiveness analysis to evaluate costs and effect of treatment versus no treatment for mild GDM using the IADPSG criteria.\textsuperscript{22} The primary outcome was incremental costs per quality-adjusted life year (QALY). They found that incremental costs per QALY were $20 412 and this was considered to be cost-effective (below cost-effectiveness threshold of $100 000/QALY).

We did not explore patient preferences. Although the OGGT in general is considered to be an unpleasant test for pregnant women,\textsuperscript{25,26} we do not have information on women’s preferences regarding the OGGT and other (screening) tests.

Before publication of the guideline on “Diabetes and Pregnancy” by the Dutch Society of Obstetrics and Gynaecology in 2010, many different tests for GDM were used. The test most frequently used to perform screening was random glucose testing. The test most frequently used to diagnose GDM was a “breakfast” or “lunch” test (43%). Based on the results presented in this thesis random glucose testing should not be used in the second trimester to screen for GDM. Diagnosing GDM should be done with an OGGT instead of a “breakfast” or “lunch” test because for the latter associations with perinatal outcome are unclear. In line with the results presented in this thesis the guideline published in 2010 recommends the 75-g OGGT for diagnosis and does not recommend random glucose testing in the 2nd trimester. In view of the treatment effect of the two randomised clinical trials and in view of the results of our cost-effectiveness analysis, one could consider a routine screen test, being a fasting glucose or an OGGT for all women.

Implications for further research

In discussing screening for GDM it is important to keep in mind the purpose of our actions. By means of screening we intend to select women with GDM in order to offer them treatment, with the aim of to prevent adverse pregnancy outcomes. Therefore we need an accurate test or testing strategy that can provide cost-effective screening with preferably minimal burden to women subjected to the test. New criteria for GDM have been proposed by the IADPSG and if these criteria are going to be accepted internationally, they need to be evaluated for their effect on health care economics and
pregnancy outcomes. Furthermore the effect of treatment should be evaluated in the light of these new criteria and possibilities of screening strategies should be explored.

Based on the findings in this thesis there seems to be a role for risk factor based screening, although refining this role by means of validation of the model presented in this thesis or development of new models is essential and should preferably be done in large observational studies. From our cost effective analysis we conclude that incorporation of patient characteristics in a screening strategy for GDM does reduce cost, but only at the expense of a relative high rate of GDM related perinatal complications. Therefore, at this point we recommend either a fasting glucose test or an OGTT for all pregnant women. Possibilities of combining results of a prediction model with fasting glucose measurement should be further explored.
References


