Screening for Gestational Diabetes Mellitus

Marsha van Leeuwen

Screening for Gestational Diabetes Mellitus
Thesis, University of Amsterdam, the Netherlands

The printing of this thesis was supported by: Diabetes Fonds, Goodlife Healthcare, Medical Dynamics, BMA BV (Mosos), Boehringer Ingelheim BV, Abbott Diabetes Care, Memidis Pharma BV, Posthumus Meyjesfonds, Vrouwenkliniek AMC.

Cover	Jana van der Merwe
Lay-out	Chris Bor
Printed by	Buijten & Schipperheijn, Amsterdam
ISBN	978-90-9027056-2

Screening for Gestational Diabetes Mellitus

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college
voor promoties ingestelde commissie,
in het openbaar te verdedigen
in de Agnietenkapel op

woensdag 17 oktober 2012,

te 14:00 uur

door

Marsha van Leeuwen

geboren te Den Haag

Promotiecommissie

Promotores	Prof. dr. B.W.J. Mol Prof. dr. G.H.A. Visser
Co-promotor	Dr. B.C. Opmeer
Overige leden	Prof. dr. J.A.M. van der Post Prof. dr. H.P. Sauerwein Dr. I.M. Evers Dr. G. ter Riet Prof. dr. A. Franx

Faculteit der Geneeskunde

Aan mijn ouders

Contents

Chapter 1

General introduction

Gestational diabetes mellitus: a metabolic disorder of pregnancy

Gestational diabetes mellitus (GDM) is a common metabolic disorder of pregnancy. It is defined as carbohydrate intolerance of variable severity with onset or first recognition during pregnancy.[1] It includes new-onset carbohydrate intolerance as well as pre-existing diabetes mellitus that has not been recognised before pregnancy (predominantly diabetes mellitus type 2).[1] GDM comprises carbohydrate intolerance that continues to exist after pregnancy as well as carbohydrate intolerance that subsides after pregnancy.

Pathophysiology

Placental hormones produced in pregnancy hamper normal carbohydrate metabolism. Corticotropin-releasing hormone, progesterone, oestradiol, and human placental lactogen interfere (directly or indirectly) with insulin receptors that are situated on various cells of the human body, making these cells less sensitive for insulin. Decreased insulin sensitivity leads to diminished entry of glucose into the cells and thus hyperglycemia. This is a physiological process emerging in the second trimester of pregnancy, enabling the growing fetus adequate glucose supply. To maintain maternal blood glucose levels within the normal range, production of insulin by the beta-cells of the pancreas is increased.[2] If this mechanism is insufficient, GDM may occur. It is unclear why some women develop GDM while others do not. In non-pregnant obese individuals, elevated levels of free fatty acids lead to decreased insulin sensitivity. In case of pregnancy, insulin sensitivity decreases even further, potentially leading to hyperglycemia.[3,4] Another suggested mechanism is autoimmunity. A small subgroup of women with GDM carries markers of humoral autoimmunity against pancreatic beta-cells. Autoimmunity against pancreatic beta-cells in otherwise asymptomatic women may become manifest in pregnancy as insulin resistance is increased due to interference of placental hormones with the insulin receptors.[5]

Criteria for GDM

Over the years there have been many different criteria for GDM all reflecting carbohydrate intolerance, albeit at different levels.[1,6] Original criteria for GDM were established by O'Sullivan and Mahan in 1964.[7] These criteria were initially selected to identify women at risk for developing diabetes mellitus (type 2) in the future, and did not reflect the risk for complications during pregnancy and delivery. In recent years, the focus has been more directed on perinatal and short-term maternal outcomes.[8]

The reference test to diagnose GDM is the oral glucose tolerance test (OGTT). With this test a glucose solution containing 75 g or 100 g of glucose is ingested after overnight fasting. Before and after administration of the glucose containing solution, plasma glucose values are measured.[9-12] As various threshold values are applied to classify the

Table 1. Various criteria for gestational diabetes mellitus (GDM)

	OGTT	Fasting value	1 hour value	2 hour value	3 hour value
World Health Organisation [a]	75 g	> 7.0 mmol/l	-	≥ 7.8 mmol/l	-
American Diabetes Association [b]	75 g	≥ 5.3 mmol/l	≥ 10.0 mmol/l	≥ 8.6 mmol/l	-
American Diabetes Association [b]	100 g	≥ 5.3 mmol/l	≥ 10.0 mmol/l	≥ 8.6 mmol/l	≥ 7.8 mmol/l

OGTT = Oral glucose tolerance test; [a] Diagnosis of GDM if fasting value > 7.0 mmol/L or 2 hour value ≥ 7.8 mmol/L; [b] Diagnosis of GDM if two out of three of four values exceed threshold value

results of the OGTT as abnormal, (international) comparison of the prevalence of GDM is complicated.[9-12] Common criteria to define the OGTT as abnormal are criteria set by the World Health Organisation (WHO) and the American Diabetes Association (ADA) (Table 1).[1,6]

Epidemiology

The prevalence of GDM depends on the criteria that are used. Furthermore, women with specific ethnic background (e.g. women from Asia, the Caribbean, and the Middle East) are at increased risk for GDM. Probably genetic differences, access to health care and dietary habits play a role.[13] In the USA and Canada the prevalence of GDM is reported to be between 2.5 and 10%, depending on ethnicity of the population studied.[14] In a number of Asian countries (e.g. India) GDM is reported in up to 15% of all pregnancies.[15] In the Netherlands prevalence is estimated to vary between 2 and 4%. Worldwide the prevalence of GDM is rising, mainly due to the rising epidemic of overweight and obesity.[4,16,17] Other risk factors for GDM reported in the literature are: family history of diabetes mellitus, increasing maternal age, obstetric history (previous GDM or offspring with birth weight >90th percentile), multiple pregnancy and polycystic ovarian syndrome.[14]

Clinical relevance

Pre-existing diabetes mellitus type 1 and 2 are associated with maternal complications and adverse perinatal outcome (such as congenital anomalies).[18,19] When we started the studies described in this thesis, the association between GDM and the risk of pregnancy complications had been described in various studies, however it seemed less explicit than in overt diabetes. As high concentrations of glucose in women with GDM result in increased fetal insulin production, and fetal hyperinsulinemia leads amongst others to macrosomia, most reported adverse outcome associated with GDM in the literature was fetal overgrowth, with related complications such as caesarean section.[20]

In 2008 the results of the Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) study were published. In this study the association between hyperglycemia in pregnancy (less severe than pre-existing diabetes) and adverse maternal and pregnancy outcomes was evaluated.[21] Higher levels of blood glucose after a 75-g OGTT were associated

with increased risk of primary caesarean section, clinical neonatal hypoglycemia, preterm delivery, preeclampsia and shoulder dystocia.[21] Women with GDM also have an increased risk of developing GDM in their next pregnancy. Moreover, the risk of developing diabetes type 2 in the future is increased. This risk may be as high as 50%, and is highest in women who need insulin treatment in pregnancy, and in women with obesity. [22] Children of women with GDM have an increased risk for childhood and adult metabolic problems, such as obesity and diabetes mellitus type 2.[23]

Effect of treatment

Results of two large randomised controlled trials have shown that treatment of GDM with dietary and life style advices, and insulin if required, is associated with a significant reduction of pregnancy complications compared to routine obstetric care, indicating a beneficial effect of treatment of GDM. [24-26]

Background of the research described in this thesis

When we started the studies described in this thesis, there was considerable debate on the relevance of screening for GDM. Internationally but also nationwide various criteria for GDM were used and different strategies for the diagnostic work-up were applied. This was mainly due to lack of evidence on the association between hyperglycemia in pregnancy and pregnancy complications. The beneficial effect of treatment at the time was established in one randomised controlled trial.[24] Throughout the period this thesis was written, the association between maternal hyperglycemia and pregnancy complications became clear from the HAPO study.[21] Also, results of another randomised controlled trial on the effect of treatment of GDM were published.[25] Although the need for a uniform diagnostic strategy was there for a long time, results of the above mentioned studies emphasised this need even more.

The OGTT can be used as a diagnostic test for GDM in women with symptoms, e.g. suspected macrosomia or polyhydramnios. Often women with GDM however, have no specific symptoms or signs. In asymptomatic women, GDM can only be detected if some form of glucose screening is performed. In current clinical practice various tests and strategies are used. Frequently used tests are random glucose measurement, fasting glucose measurement and a glucose challenge test (blood glucose measurement one hour after ingestion of 50 g of glucose).[27] There is no agreement on which screening test is most appropriate, since estimates of accuracy and costs of the tests reported in the literature vary. There is also discussion on which women should be tested. International expert groups have recommended selective screening based on risk factors for GDM.[28] Opponents of this selective screening strategy criticise the use of risk factors, since this strategy fails to identify over one-third of cases of GDM, and therefore they advocate universal screening.[1,29-31] Studies that have evaluated risk factor based screening often included a selection of risk factors and did not combine these risk factors in a quantitative

manner (i.e. risk scoring system or a prediction model), thereby possibly overlooking the diagnostic value of risk factors in the selection of women at risk for GDM.

Summary of background

Worldwide the prevalence of GDM is rising. Hyperglycemia in pregnancy is associated with maternal and pregnancy complications. Treatment of GDM reduces the risk of pregnancy complications. The question remains what is the best strategy to identify women with GDM?

Aim of this thesis

The aim of this thesis was to evaluate various screening strategies for GDM. We wanted to explore accuracy measures of three individual screening tests, and the potential of using risk factors and patient characteristics. Furthermore, we wanted to assess costs associated with the various strategies, in order to obtain an adequate and cost-effective strategy to timely detect women with GDM.

Specific research questions were

1. What is the current policy on the (diagnostic) work-up of GDM in the Netherlands?
2. Which test has the best accuracy to screen for GDM?
3. Is it possible to estimate the risk of GDM for individual patients?
4. Can we use risk indicators to improve (cost-) effectiveness or of screening?
5. What is the most cost-effective strategy to prevent complications from GDM?

Outline of the thesis

This thesis comprises three parts: In the first part we assess the accuracy of three different screening tests for GDM. In the second part we focus on the use of individual patient characteristics and on risk factors to predict GDM. In the third part clinical practice in the Netherlands and costs and effects of various screening strategies are evaluated.

In **chapter 2** we report the results of a comparison of the performance of two screening tests for GDM. Accuracy measures of the random glucose test and the 50-g glucose challenge test were compared with data from a prospective cohort of 1301 women. All women underwent a 50-g glucose challenge test as well as random glucose measurement between 24 and 28 weeks of gestation with the 75-g OGTT as gold standard. We assessed the performance of the tests by their discriminative capacities (Q2).

In **chapter 3** we describe the results of a systematic review of the literature to assess the accuracy of the random glucose test as a screening test for GDM (Q2).

In **chapter 4** we describe the results of a systematic review of the literature and meta-analysis to obtain summary estimates of accuracy measures of the 50-g glucose challenge test for GDM using a bivariate approach (Q2).

In **chapter 5** we report on the accuracy of fasting glucose measurement for the detection of GDM. With a systematic review of the literature and bivariate meta-analysis we estimated summary estimates of accuracy measures and assessed the ability of the test as screening test for GDM (Q2).

In **chapter 6** we present results of a validation study in which we evaluated a clinical scoring system for GDM. To validate the scoring system we used data from a prospective cohort study comprising 1266 women. Performance of the scoring system was evaluated in terms of calibration and discriminative ability. We compared the efficiency of a screening strategy derived from the scoring system with conventional screening (Q3 and Q4).

In **chapter 7** we describe the construction and internal validation of a clinical prediction model based on medical history and patient characteristics to estimate the risk of GDM in individual women. We constructed the prediction model with multiple logistic regression analysis with data from a prospective cohort study and evaluated its performance with internal validation (Q3 and Q4).

In **chapter 8** we give an overview of the literature on the increased risk of perinatal complications associated with GDM, and on the effect of treatment of GDM. In this chapter we also report the results of a survey among gynaecologists and midwives to assess current policy and clinical practice regarding detection and treatment of GDM in the Netherlands (Q1).

In **chapter 9** costs and effects of various strategies to detect GDM are evaluated in a model based economic evaluation. A detailed cost-effective analysis was performed calculating incremental costs per prevented perinatal complication (Q5).

In **chapter 10** we summarise and discuss the results of this thesis and evaluate their implications for clinical practice and for future research.

In **chapter 11** we summarise and discuss the results of this thesis and evaluate their implications for clinical practice and for future research in Dutch.

References

1. American Diabetes Association. Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 2009; 32(Supp 1):S62-S67.
2. Kühl C. Etiology and pathogenesis of gestational diabetes. Diabetes Care 1998; 21(Suppl 2):B19-26.
3. Tataranni P. Pathophysiology of obesity-induced insulin resistance and type 2 diabetes mellitus. Eur Rev Med Pharmacol Sci 2002; 6:27-32.
4. Yogev Y, Visser GH. Obesity, gestational diabetes and pregnancy outcome. Semin Fetal Neonatal Med 2009; 14:77-84.
5. De Leiva A, Mauricio D, Corcoy R. Diabetes-related autoantibodies and gestational diabetes. Diabetes Care 2007; 30(S2):S127-133.
6. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and Classification of Diabetes Mellitus; Report of a WHO consultation. Diabet Med 1998; 15(7):539-53.
7. O'Sullivan JB, Mahan C. Criteria for oral glucose tolerance test in pregnancy. Diabetes 1964; 13:278-285.
8. Jensen DM, Korsholm L, Ovesen P, Beck-Nielsen H, Molsted-Pedersen L, Damm P. Adverse pregnancy outcome in women with mild glucose intolerance: is there a clinically meaningful threshold value for glucose? Acta Obstet Gynecol Scand 2008; 87(1):59-62.
9. National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. Diabetes 1979; (28):1039-1057.
10. Lind T, Phillips PR. The Diabetic Pregnancy Study Group of the European Association for the Study of Diabetes. Influence of pregnancy on the 75-g OGTT. A prospective multicenter study. Diabetes 1991; 40(Suppl 2):8-13.
11. Hoffman L, Nolan C, Wilson JD. The Australasian Diabetes in Pregnancy Society. Gestational diabetes mellitus, management guidelines. Med J Aust 1998; 169:93-97.
12. American Diabetes Association. American Diabetes Association clinical practice recommendations 2001. Diabetes Care 2001; 24(Suppl 1): S1-S133.
13. De Silva NM, Frayling TM. Novel biological insights emerging from genetic studies of type 2 diabetes and related metabolic traits. Curr Opin Lipidol 2010; 21(1):44-50.
14. King H. Epidemiology of glucose intolerance and gestational diabetes in women of childbearing age. Diabetes Care 1998; 21(Suppl 2):B9-13.
15. Seshiah V, Balaji V, Balaji MS, Paneerselvam A, Arthi T, Thamizharasi M, et al. Prevalence of gestational diabetes mellitus in South India (Tamil Nadu)-a community based study. J Assoc Physicians India 2008; 56:329-33.
16. Getahun D, Nath C, Ananth CV, Chavez MR, Smulian JC. Gestational diabetes in the United States: temporal trends 1989 through 2004. Am J Obstet Gynecol 2008; 198(5):525.e1-5.
17. Hunt KJ, Schuller KL. The increasing prevalence of diabetes in pregnancy. Obstet Gynecol Clin North Am 2007; 34(2):173-99.
18. Evers IM, de Valk HW, Visser GH. Risk of complications of pregnancy in women with type 1 diabetes: nationwide prospective study in the Netherlands. BMJ 2004; 328(7445):915-921.
19. Macintosh MC, Fleming KM, Bailey JA, Doyle P, Modder J, Acolet D, et al. Perinatal mortality and congenital anomalies in babies of women with type 1 or type 2 diabetes in England, Wales, and Northern Ireland: population based study. BMJ 2006; 333(7560):177-183.
20. Catalano PM, Hauguel-De Mouzon S. Is it time to revisit the Pedersen hypothesis in the face of the obesity epidemic? Am J Obstet Gynecol 2011; 204(6):479-87.
21. Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR, et al. HAPO Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008; 358(19):1991-2002.

22. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. Diabetes Care 2002; 25(10):1862-8.
23. Plagemann A, Roepke K, Harder T, Brunn M, Harder A, Wittrock-Staar M, et al. Epigenetic malprogramming of the insulin receptor promoter due to developmental overfeeding. J Perinat Med 2010; 38(4):393-400.
24. Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS; Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) Trial Group. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med 2005; 352(24):2477-86.
25. Landon MB, Spong CY, Thom E, Carpenter MW, Ramin SM, Casey B, et al. A multicenter, randomized trial of treatment for mild gestational diabetes. N Engl J Med 2009; 361(14):1339-48.
26. Horvath K, Koch K, Jeitler K, Matyas E, Bender R, Bastian H, et al. Effects of treatment in women with gestational diabetes mellitus: systematic review and meta-analysis. BMJ 2010; 340:c1395.
27. Griffin ME, Coffey M, Johnson H, Scanlon P, Foley M, Stronge J, et al. Universal vs. risk factor-based screening for gestational diabetes mellitus: detection rates, gestation at diagnosis and outcome. Diabet Med 2000; 17:26-32.
28. Teh WT, Teede HJ, Paul E, Harrison CL, Wallace EM, Allan C. Risk factors for gestational diabetes mellitus: implications for the application of screening guidelines. Aust N Z J Obstet Gynaecol 2011; 51(1):26-30.
29. Alberico S, Strazzanti C, De Santo D, De Seta F, Lenardon P, Bernardon M, et al. Gestational diabetes: universal or selective screening? J Matern Fetal Neonatal Med 2004; 16:331-7.
30. Cosson E, Benchimol M, Carbillon L, Pharisien I, Paries J, Valensi P, et al. Universal rather than selective screening for gestational diabetes mellitus may improve fetal outcomes. Diabetes Metab 2006; 32:140-6.
31. Virally M, Laloi-Michelin M. Methods for the screening and diagnosis of gestational diabetes mellitus between 24 and 28 weeks of pregnancy. Diabetes Metab 2010; 36(6 Pt 2):549-65.

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.[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.[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.[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.[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.

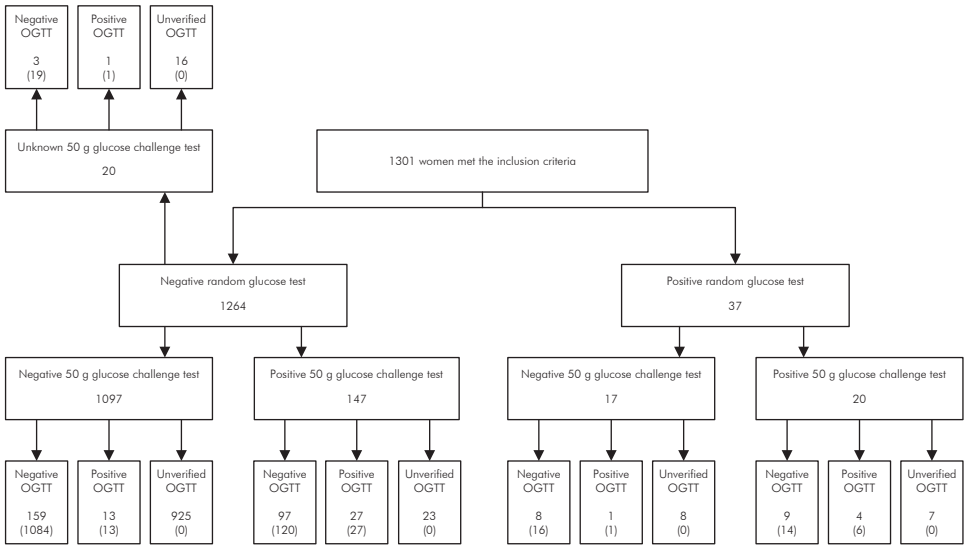
Table 1. Demographics before correction for verification bias.

	GDM present (N = 46)	GDM not present (N = 276)	GDM not verified (N = 979)	Total (N = 1301)
Age (years)	30.8 ± 4.6	30.6 ± 4.9	30.8 ± 5.0	30.8 ± 4.9
BMI before pregnancy (kg/m^2)	25.6 ± 4.4	25.7 ± 5.2	23.8 ± 4.4	24.2 ± 4.6
Ethnicity				
Caucasian	37 (82.2)	247 (90.5)	848 (89.4)	1132 (89.4)
Non Caucasian	8 (17.8)	26 (9.5)	101 (10.6)	135 (10.6)
Family history of diabetes				
Yes	13 (28.9)	55 (20.1)	185 (19.5)	253 (19.7)
No	32 (71.1)	218 (79.9)	783 (80.5)	1033 (80.3)
Smoking				
Yes	8 (17.4)	46 (17.0)	170 (17.5)	224 (17.4)
No	38 (82.6)	225 (83.0)	799 (82.5)	1062 (82.6)
Hospital				
Utrecht	22 (47.8)	99 (35.9)	874 (89.3)	995 (76.5)
Zwolle	24 (52.2)	177 (64.1)	105 (10.7)	306 (23.5)
Obstetric history 1				
Previous miscarriage	15 (32.6)	84 (30.4)	317 (32.4)	416 (32.0)
No previous miscarriage	31 (67.4)	192 (69.6)	662 (67.6)	885 (68.0)
Obstetric history 2				
Nullipara	19 (43.2)	112 (40.9)	422 (44.3)	553 (43.5)
Multipara with history of GDM	2 (4.5)	8 (2.9)	6 (0.6)	16 (1.3)
Multipara without history of GDM	23 (52.3)	154 (56.2)	525 (55.1)	702 (55.2)
Obstetric history 3				
Nullipara	19 (43.2)	112 (40.7)	422 (44.3)	553 (43.5)
Multipara with perinatal mortality	4 (9.1)	17 (6.2)	47 (4.9)	68 (5.3)
Multipara without perinatal mortality	21 (47.7)	146 (53.1)	484 (50.8)	651 (51.2)

Data are means ± SD or n (%).

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

Figure 1. Screening test and OGTT results before and after correction for verification bias.



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.

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

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.

	Random glucose test			1 hour 50-g glucose challenge test		
	OGTT abnormal	OGTT normal	Total	OGTT abnormal	OGTT normal	Total
Screening test abnormal	7	30	37	33	134	167
Screening test normal	41	1223	1264	14	1100	1114
Total	48	1253	1301	47	1234	1281
Sensitivity (%)	14.6 (4.6-24.6)			70.2 (57.1-83.3)		
Specificity (%)	97.6 (96.6-98.5)			89.1 (87.4-90.9)		
Positive predictive value (%)	18.9 (6.3-31.5)			19.8 (3.7-25.8)		
Negative predictive value (%)	96.8 (91.0-100.0)			98.7 (97.1-100.0)		
Likelihood Ratio abnormal test result	6.1 (2.8-13.2)			6.5 (5.1-8.3)		
Likelihood Ratio normal test result	0.88 (0.78-0.98)			0.33 (0.22-0.52)		
Diagnostic Odds Ratio	7.0 (2.9-16.8)			19.4 (6.8-31.9)		
Area under the curve	0.69 (0.61-0.78)			0.88 (0.83-0.93)		

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.[12,16] 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.[17]

There are several strategies to deal with incomplete data, also within the context of partial verification.[17] 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%.[1] 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.[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 26th to 30th week of gestation.[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.[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%.[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.[26] Jowett *et al*. also concluded that random glucose measurement is not sufficiently sensitive for screening on GDM.[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

1. Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med 2005; 352:2477-2486.
2. Esakoff TF, Cheng YW, Caughey AB. Screening for gestational diabetes: different cut-offs for different ethnicities? Am J Obstet Gynecol 2005; 193:1040-1044.
3. Jensen DM, Damm P, Sorensen B, Molsted-Pedersen L, Westergaard JG, Korsholm L et al. Proposed diagnostic thresholds for gestational diabetes mellitus according to a 75-g oral glucose tolerance test. Maternal and perinatal outcomes in 3260 Danish women. Diabet Med 2003; 20:51-57.
4. Sivan E, Weisz B, Homko CJ, Reece EA, Schiff E. One or two hours postprandial glucose measurements: are they the same? Am J Obstet Gynecol 2001; 185:604-607.
5. Tuffnell DJ, West J, Walkinshaw SA. Treatments for gestational diabetes and impaired glucose tolerance in pregnancy. Cochrane Database Syst Rev 2003 CD003395.
6. Brody SC, Harris R, Lohr K. Screening for gestational diabetes: a summary of the evidence for the U.S. Preventive Services Task Force. Obstet Gynecol 2003; 101:380-392.
7. Langer O, Yogev Y, Most O, Xenakis EM. Gestational diabetes: the consequences of not treating. Am J Obstet Gynecol 2005; 192:989-997.
8. Naylor CD, Sermer M, Chen E, Farine D (Toronto Trihospital Gestational Diabetes Project Investigators). Selective screening for gestational diabetes mellitus. N Engl J Med 1997; 337:1591-1596.
9. Poyhonen-Alho MK, Teramo KA, Kaaja RJ, Hiilesmaa VK. 50-gram oral glucose challenge test combined with risk factor-based screening for gestational diabetes. Eur J Obstet Gynecol Reprod Biol 2005; 121:34-37.
10. Santos-Ayarzagoitia M, Salinas-Martinez AM, Villarreal-Perez JZ. Gestational diabetes: Validity of ADA and WHO diagnostic criteria using NDDG as the reference test. Diabetes Res Clin Pract 2006; 74:322-328.
11. American Diabetes Association. Gestational diabetes mellitus. Diabetes Care 2003; 26 Suppl 1:S103-S105.
12. Begg CB. Biases in the assessment of diagnostic tests. Stat Med 1987; 6:411-423.
13. Rubin DB. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons Inc. New York, 1987.
14. Schafer JL. Analysis of Incomplete Multivariate Data. Chapman and Hall. New York, 1997.
15. Harrell F. Regression Modeling Strategies. Springer. New York, 2001.
16. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 1983; 39:207-215.
17. Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press. Oxford, 2003.
18. Hunink MG, Richardson DK, Doubilet PM, Begg CB. Testing for fetal pulmonary maturity: ROC analysis involving covariates, verification bias, and combination testing. Med Decis Making 1990; 10:201-211.
19. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. Biostatistics 2001; 2:249-260.
20. Mathai M, Thomas TJ, Kuruvila S, Jairaj P. Random plasma glucose and the glucose challenge test in pregnancy. Natl Med J India 1994; 7:160-162.
21. McElduff A, Goldring J, Gordon P, Wyndham L. A direct comparison of the measurement of a random plasma glucose and a post-50 g glucose load glucose, in the detection of gestational diabetes. Aust N Z J Obstet Gynaecol 1994; 34:28-30.

22. Perucchini D, Fischer U, Spinas GA, Huch R, Huch A, Lehmann R. Using fasting plasma glucose concentrations to screen for gestational diabetes mellitus: prospective population based study. BMJ 1999; 319:812-815.
23. Rey E, Hudon L, Michon N, Boucher P, Ethier J, Saint-Louis P. Fasting plasma glucose versus glucose challenge test: screening for gestational diabetes and cost effectiveness. Clin Biochem 2004; 37:780-784.
24. de Sereday MS, Damiano MM, Gonzalez CD, Bennett PH. Diagnostic criteria for gestational diabetes in relation to pregnancy outcome. J Diabetes Complications 2003; 17:115-119.
25. Maresh M. Screening for gestational diabetes mellitus. Semin Fetal Neonatal Med 2005; 10:317-323.
26. Nasrat AA, Johnstone FD, Hasan SA. Is random plasma glucose an efficient screening test for abnormal glucose tolerance in pregnancy? Br J Obstet Gynaecol 1988; 95:855-860.
27. Jowett NI, Samanta AK, Burden AC. Screening for diabetes in pregnancy: is a random blood glucose enough? Diabet Med 1987; 4:160-163.

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.[1] Treatment of GDM improves perinatal as well as maternal outcome.[2,3] 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.[4] Currently there is no consensus on the optimal approach to screen for GDM.[4,5] 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.[6-9] 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.[10] 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.[11] 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.[12-14] 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.[15] No methodological filter for diagnostic accuracy studies or any other restriction was applied as this can lead to omission of relevant papers.[15,16] 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.[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.[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.[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.[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

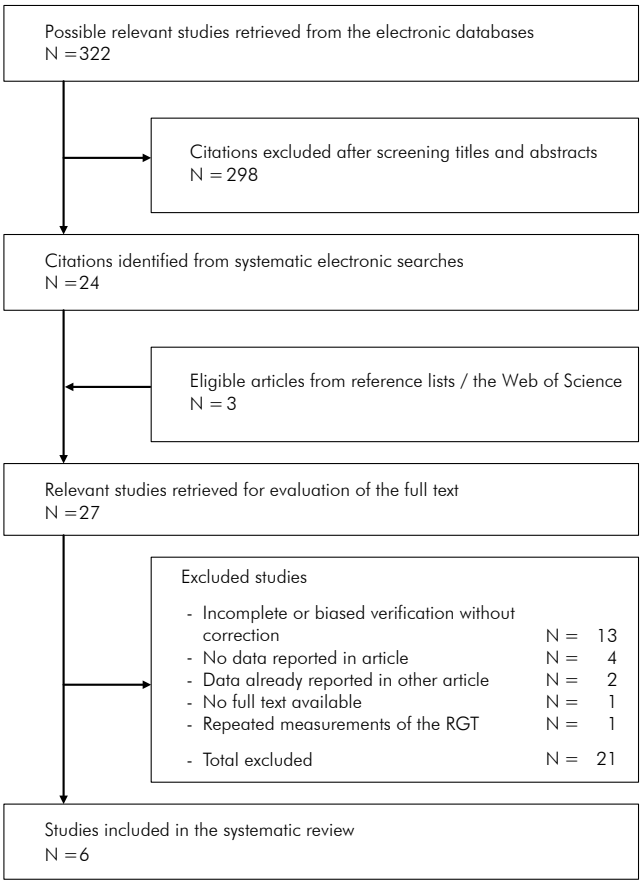


Figure 1. Results of the literature search and study selection.

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.[13,14,20,21] In two studies multiple RGTs were performed.[12,22] 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

GDM. [13,14,20,21] 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.[20] 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%.[13]

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.[12] 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. [21] 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).[22]

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.[23-25] We were unable to obtain the full text of one of these studies.[23] 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

First author, Year	Country	N	Inclusion criteria	Gestational age (wks)	Verification OGTT (%)	GDM n (%)
Jowett [22, a] 1987	UK	110	Risk factors	27 - 31	100	49 (45)
Nasrat [20] 1988	Kuwait	276 (250)	All women	28 - 32	91	50 (20)
Mathai [13] 1994	India	232	All women	26 - 30	100	11 (4.7)
Tam [21] 2000	China	895	All women	24 - 32	100	108 (12)
Maegawa [12, a] 2003	Japan	749 (735 [b])	All women	1st & 2nd trimester	100	1.9/1.1 [b] (14/8 [b])
Van Leeuwen [14] 2007	The Netherlands	1301	All women	24 - 28	100 [c]	48 (3.8)

[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.[26] If publication bias is present, the accuracy of the RGT reported in this systematic review was most likely overestimated.

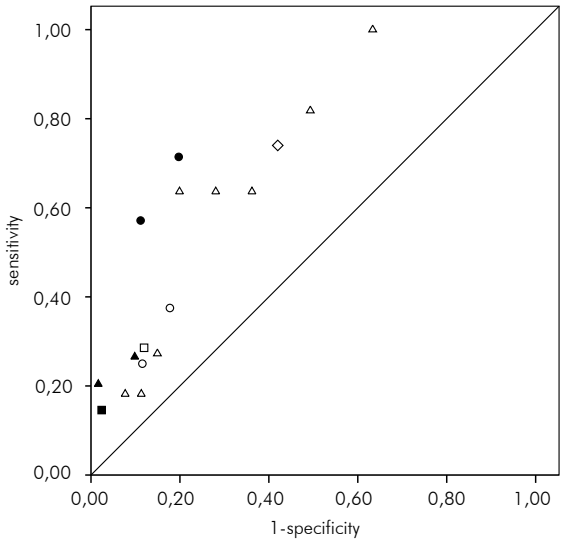


Figure 2. ROC plot of the included studies for various thresholds the RGT. Displayed are the studies of Mathai (Δ), Nasrat (□), Tam (◊), van Leeuwen *et al*. (■), Maegawa *et al*. (●, first trimester and ○, second trimester), Jowett (▲).

OGTT (gram)	Cut off OGTT (mmol/L)	Time RGT - OGTT	Time to last meal	Blood glucose	Cut off RGT (mmol/L)
75	8.0 [f]	Same day	Fixed times	Venous plasma	5.6 and 6.1
75	8.0 [f]	< 5 days	Independent	Venous plasma	5.8 and 6.9
100	5.3; 10; 8.6; and 7.8 [e]	< 2 weeks	Independent	Venous plasma	4.4 to 6.4
75	8.0 [f]	< 4 weeks	Independent	Venous plasma	4.7
75	5.6; 10; and 8.3 [d]	2 to 4 weeks	Independent	Venous plasma	5.3 and 5.6
75	7.8 [g]	< 2 weeks	Independent	Venous plasma	6.8

[e]. Fasting glucose value, glucose value at 1, 2, and 3 hours after the glucose load (GDM was diagnosed if 2 values exceeded threshold values); [f]. 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); [g]. Two hour glucose value

Table 2. Study quality per study of the six included studies assessed with the QUADAS list.

	Jowett [22]	Nasrat [20]	Mathai [13]	Tam [21]	Maegawa [12]	van Leeuwen [14]
Patients representative of practice	No	Yes	Yes	Yes	Yes	Yes
Clear description selection criteria	Yes	Yes	Yes	Yes	Yes	Yes
Reference test likely to detect GDM	Yes	Yes	Yes	Yes	Yes	Yes
Time between tests short enough	Yes	Yes	Yes	Yes	Yes	Yes
Complete verification index test	Yes	Yes	Yes	Yes	Unknown	Yes
Consistent performance reference test	Yes	Yes	Yes	Yes	Yes	Yes
Index and reference test performed independently	Yes	Yes	Yes	Yes	Yes	Yes
Clear description of index test	Yes	Yes	Yes	No	No	Yes
Clear description of reference test	Yes	Yes	Yes	No	No	Yes
RGT interpreted without results OGTT	Yes	Unknown	Unknown	Unknown	Unknown	Yes
OGTT interpreted without results RGT	No	No	Unknown	Unknown	Unknown	No
Clinical data same as practice	Yes	Yes	Yes	Unknown	No	Yes
Unintepretable test results reported	Unknown	Yes	Yes	Yes	Unknown	Yes
Withdrawals explained	Unknown	Yes	Yes	Yes	Unknown	Yes
Intervention between index and reference test	No	No	No	No	No	No

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.[8] 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.[11] 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.[12] 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.[27] 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.[28] 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.[29] 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

1. Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR, *et al*. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008; 358(19):1991-2002.
2. Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med 2005; 352(24):2477-86.
3. Landon MB, Shriver EK. A prospective multicenter randomized treatment trial of mild gestational diabetes (GDM) (Abstract). Am J Obstet Gynecol 2008; 199(6):S2.
4. Hollander MH, Paarlberg KM, Huisjes AJ. Gestational diabetes: a review of the current literature and guidelines. Obstet Gynecol Surv 2007; 62(2):125-36.
5. Brown CJ, Dawson A, Dodds R, Gamsu H, Gillmer M, Hall M, *et al*. Report of the Pregnancy and Neonatal Care Group. Diabet Med 1996; 13(9 Suppl 4):S43-S53.
6. American Diabetes Association. Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 2009; 32: Suppl. 1:S62-7.
7. Canadian Diabetes Association. Canadian Diabetes Association. Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. Canadian Diabetes Association. http://www.diabetes.ca/files/cpg2008/cpg-2008.pdf (2008).
8. National Collaborating Centre for Women's and Children's Health. Diabetes in pregnancy. Management of diabetes and its complications from preconception to the postnatal period. http://www.ncc-wch.org.uk/index.asp?PageID=343 (2008).
9. Hoffman L, Nolan C, Wilson JD, Oats JJ, Simmons D. Gestational diabetes mellitus-management guidelines. The Australasian Diabetes in Pregnancy Society. Med J Aust 1998; 169(2):93-7.
10. Hanna FW, Peters JR, Harlow J, Jones PW. Gestational diabetes screening and glycaemic management; national survey on behalf of the Association of British Clinical Diabetologists. QJM 2008; 101(10):777-84.
11. van Leeuwen M, Prins SM, de Valk HW, Evers IM, Visser GH, Mol BW. Gestational diabetes mellitus: treatment reduces the risk of complications. Ned Tijdschr Geneeskd. 2011; 155:A2291
12. Maegawa Y, Sugiyama T, Kusaka H, Mitao M, Toyoda N. Screening tests for gestational diabetes in Japan in the 1st and 2nd trimester of pregnancy. Diabetes Res Clin Pract 2003; 62(1):47-53.
13. Mathai M, Thomas TJ, Kuruvila S, Jairaj P. Random plasma glucose and the glucose challenge test in pregnancy. Natl Med J India 1994; 7(4):160-2.
14. van Leeuwen M, Zweers EJ, Opmeer BC, van Ballegooie E, ter Brugge HG, de Valk HW, *et al*. Comparison of accuracy measures of two screening tests for gestational diabetes mellitus. Diabetes Care 2007; 30(11):2779-84.
15. de Vet HCW, Eisinga A, Riphagen II, Aertgeerts B, Pewsner D. Chapter 7: Searching for Studies. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4 [updated September 2008]. The Cochrane Collaboration, 2008. http://srdta.cochrane.org/Files/Website/Handbook/Chapter07-Searching-(September-2008).pdf.
16. Leeflang MM, Scholten RJ, Rutjes AW, Reitsma JB, Bossuyt PM. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. J Clin Epidemiol 2006; 59(3):234-40.
17. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol 2006; 6:9.
18. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005; 58(10):982-90.
19. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and Classification of Diabetes Mellitus; Report of a WHO consultation. 1999.

20. Nasrat AA, Johnstone FD, Hasan SA. Is random plasma glucose an efficient screening test for abnormal glucose tolerance in pregnancy? Br J Obstet Gynaecol 1988; 95(9):855-60.
21. 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.
22. Jowett NI, Samanta AK, Burden AC. Screening for diabetes in pregnancy: is a random blood glucose enough? Diabet Med 1987; 4(2):160-3.
23. Kurishita M, Tateyama Y, Itasaka T, Hayashi M, Hosono Y, Itoh H, *et al*. Utility of measurement of random blood sugar levels for screening pregnant women who have glucose intolerance. Rinsyo Fujinka Sanka 1997; 51:337-41.
24. Maheshwari JR, Mataliya MV, Patil. Screening for glucose tolerance in pregnancy utilizing random plasma glucose assay. J Obstet Gynaecol India 1989; 39:351-3.
25. Seshiah V, Hariharan RS, Sundaram A, Ramanakumar TK, Ganesan VS. "Spot test" for blood glucose in pregnancy. J Obstet Gynaecol India 1986; 36:287-91.
26. Sackett DL. Bias in analytic research. J Chronic Dis 1979; 32(1-2):51-63.
27. Chen R, Yogev Y, Ben-Haroush A, Jovanovic L, Hod M, Phillip M. Continuous glucose monitoring for the evaluation and improved control of gestational diabetes mellitus. J Matern Fetal Neonatal Med 2003; 14(4):256-60.
28. Combs CA, Gunderson E, Kitzmiller JL, Gavin LA, Main EK. Relationship of fetal macrosomia to maternal postprandial glucose control during pregnancy. Diabetes Care 1992; 15(10):1251-7.
29. Ostlund I, Hanson U. Repeated random blood glucose measurements as universal screening test for gestational diabetes mellitus. Acta Obstet Gynecol Scand 2004; 83(1):46-51.

Chapter 4

Glucose challenge test for detecting gestational diabetes mellitus

A systematic review

Marsha van Leeuwen
Marjoleine D. Louwerse
Brent C. Opmeer,
Jacqueline Limpens
Mireille J. Serlie
Johannes B. Reitsma
Ben Willem J. Mol

British Journal of Obstetrics and Gynaecology. 2012; 119:393-401

Abstract

BACKGROUND The best strategy to identify women with gestational diabetes mellitus (GDM) is unclear.

OBJECTIVE To perform a systematic review to calculate summary estimates of sensitivity and specificity of the 50-g glucose challenge test for GDM.

SEARCH STRATEGY Systematic search of MEDLINE, EMBASE and Web of Science.

SELECTION CRITERIA Articles that compared the 50-g glucose challenge test with the 75- or 100-g oral glucose tolerance test (OGTT, reference standard) before 32 weeks of gestation.

DATA COLLECTION AND ANALYSIS Two reviewers independently selected articles. Summary estimates of sensitivity and specificity, with 95% confidence intervals and summary receiver operating characteristic curves were calculated using bivariate random-effects models.

MAIN RESULTS Twenty-six studies were included (13 564 women). Studies that included women with risk factors showed a pooled sensitivity of the 50-g glucose challenge test of 0.74 (95% CI 0.62 - 0.87), a pooled specificity of 0.77 (95% CI 0.66 - 0.89) (threshold value of 7.8 mmol/l), a derived positive likelihood ratio (LR) of 3.2 (95% CI 2.0 - 5.2) and a negative LR of 0.34 (95% CI 0.22 - 0.53). In studies with consecutive recruitment, the pooled sensitivity was 0.74 (95% CI 0.62 - 0.87) for a specificity of 0.85 (95% CI 0.80 - 0.91), with a derived positive LR of 4.9 (95% CI 3.5 - 7.0) and negative LR of 0.31 (95% CI 0.20 - 0.47). Increasing the threshold for disease (OGTT result) increased the sensitivity of the challenge test, and decreased the specificity.

CONCLUSION The 50-g glucose challenge test is acceptable to screen for GDM, but cannot replace the OGTT. Further possibilities of combining the 50-g glucose challenge test with other screening strategies should be explored.

Introduction

Gestational diabetes mellitus (GDM) is a common metabolic complication of pregnancy that affects between 2% and 9% of all pregnant women in Western countries.[1-3] Hyperglycemia in pregnancy is associated with a number of adverse perinatal outcomes, such as neonatal clinical hypoglycemia, macrosomia, increased risk of shoulder dystocia, and the need for neonatal intensive care.[4] Maternal complications associated with hyperglycemia in pregnancy include an increased risk of caesarean delivery and preeclampsia.[4] Furthermore, women with GDM have up to 60% risk of developing type-2 diabetes mellitus (T2D) within 5 to 15 years of delivery,[5] and it has been suggested that children prenatally exposed to a diabetic milieu have an increased risk for the development of T2D later in life.[6,7]

Until recently there was a lack of evidence to demonstrate the beneficial effects of treatment for GDM. However, recently it has been demonstrated in a number of high-quality studies that treatment of GDM with diet or insulin reduces the risk of a number of important complications associated with GDM, thus improving both perinatal as well as maternal outcome.[3,8,9] Crowther *et al*. showed that treatment of GDM with diet or insulin significantly reduced the risk of serious perinatal complications from 4 to 1%.[3] Landon *et al*. showed that there were fewer cases of shoulder dystocia and fewer caesarean deliveries if women with mild GDM received treatment. [9] Identifying women with GDM in order to provide treatment has therefore become of eminent importance, but is difficult as clinical signs and symptoms are often absent.

Because of the lack of clinical signs and symptoms of GDM, screening tests are essential to identify women with GDM. One of the tests that is used in the diagnostic pathway is the 50-g glucose challenge test.[10] The 50-g glucose challenge test is a glucose-loading test. Women ingest a drink containing 50 g of glucose. After 1 hour the venous glucose level is measured. A 75- or 100-g diagnostic oral glucose tolerance test (OGTT) is performed when the blood glucose value is elevated after a 50-g glucose challenge test (threshold value often set as 7.2 or 7.8 mmol/l). The OGTT is a glucose-loading test in which women ingest a drink containing 75 or 100 g of glucose. The test is performed after overnight fasting. Venous glucose levels are measured before and at 1 and 2 hours after the ingestion of a glucose load. GDM is diagnosed if blood glucose values after an OGTT are elevated (multiple criteria; Table 1). A number of studies have evaluated the accuracy of the 50-g glucose challenge test as a screening test for GDM, reporting diverse results.

Although currently the 50-g glucose challenge test is not recommended in the majority of guidelines, it could be a useful test in the diagnostic work-up of GDM. The aim of this

Table 1. Criteria for high or low threshold of the oral glucose tolerance test (OGTT)

OGTT	OGTT threshold for test positivity	Threshold high or low
100 g	T0=5.8; T60=10.6; T120=9.2; T180=8.1	High
100 g	T0=5,5; T60=10,5; T120=8,9; T180=7,8	High
100 g	T0=5.3; T60=10.0; T120=8.6; T180=7.8	Low
100 g	T0=5.2; T60=8.6; T120=10.0; T180=7.8	Low
75 g	T0=5.6; T60=10.0; T120=8.3	High
75 g	T0=5.3; T60=10.6; T120=8.9	High
75 g	T120=7.8	Low

T = time in minutes after ingestion of the glucose load.

study was to systematically review and meta-analyse the accuracy of the 50-g glucose challenge test for the detection of glucose intolerance in pregnancy, in order to evaluate its applicability in the diagnostic work-up of GDM. We evaluated the applicability of the 50-g glucose challenge as a first-step screening test for GDM, and as a replacement of the current diagnostic test (OGTT).

Methods

Search strategy

A medical librarian (J.L.) undertook a systematic search of the electronic databases MEDLINE (1950 - October 2010) and EMBASE (1980 - October 2010) to identify studies reporting on the 50-g glucose challenge test in pregnant women. The search strategy consisted of free text words and subject headings (MeSH, SH) related to the target disease (GDM), population (pregnant women) and screening test (the 50-g glucose challenge test). No methodological filter or other restrictions were applied, as this can lead to the omission of relevant papers.[11] 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 imported all references into reference manager databases (Thomson ISI ResearchSoft, Carlsbad, CA, USA). Duplicate studies were excluded.

Study selection

Two reviewers independently screened the titles and abstracts of all of the studies retrieved (M.v.L. and M.D.L.). Based on the full-text manuscripts, we selected studies according to predefined inclusion criteria. Studies were included when they compared the 50-g glucose challenge test (index test) with either the 75- or the 100-g OGTT (reference standard) in pregnant women before 32 weeks of gestation, at any level of risk for GDM, and reported sufficient data to reproduce a 2x2 table from the two tests. Studies that did not report enough data for a 2x2 table, but for which data could possibly be obtained

from the authors, were also evaluated. Studies in which the OGTT was only performed in screen-positive women were excluded. We also excluded diagnostic case-control studies in which women with GDM were compared against women without GDM, as we expected overoptimistic estimates of test accuracy in these studies.[12,13] Final inclusion/exclusion decisions were made by comparison of the results of both reviewers. Disagreement was resolved by consulting a third independent reviewer (B.W.M.).

Data extraction

We extracted data on study characteristics, study quality and 2x2 tables of test accuracy. We used a pre-designed piloted data extraction form. If there were data missing on test accuracy or on other relevant characteristics, we contacted the corresponding author. Disagreement on data was resolved by discussion and consensus. If no consensus was reached, a third reviewer (B.W.M.) was consulted.

Quality assessment

The methodological quality of selected papers was evaluated using QUADAS, a tool for quality assessment of studies of diagnostic accuracy.[14] Included studies were evaluated on 15 items concerning patient selection, verification, description of the tests and description of the study population.[15]

Diagnosis GDM

The reference standard for diagnosis of GDM was the OGTT. In current clinical practice, the 75-g OGTT as well as the 100-g OGTT are used to diagnose GDM. We therefore included studies that used either the 75-g OGTT or the 100-g OGTT as reference tests. In the past, results of the OGTT were classified as normoglycemic, impaired glucose tolerance (IGT, intermediate category) or as GDM. Currently, the category intermediate IGT is not used. To facilitate comparison between studies and to enable meta-analysis, we considered women with IGT in older studies as either normoglycemic or as having GDM, according to the criteria currently used.

Data synthesis and bivariate regression model

We extracted 2x2 tables cross-classifying the results of the 50-g glucose challenge test with the results of the OGTT. We plotted their results in receiver operating characteristic plots, and created forest plots to visualise data and to explore heterogeneity. We used a bivariate regression model to calculate summary estimates of sensitivity and specificity, and their 95% confidence intervals, and to construct summary receiver operating characteristic (sROC) curves.[16] Likelihood ratios (LRs) were derived from estimates of sensitivity and specificity. In the bivariate regression approach pairs of sensitivity and specificity are jointly analysed within a single model using a random effects approach to account for variation beyond chance. In addition to chance variation and differences in thresholds, the heterogeneity in results between studies can result from bias arising from flawed design or a variation in accuracy between different clinical subgroups. To explore

these other sources of heterogeneity, the bivariate regression approach can be extended with covariates to examine whether they have an effect on sensitivity, specificity or both. We examined the following covariates for their effect on test accuracy: reference test (75- or 100-g OGTT), risk level of women in the study (consecutive inclusion versus inclusion of women with risk factors). As multiple criteria for an abnormal OGTT exist for the 75-g OGTT as well as for the 100-g OGTT, we categorised the threshold values of the OGTT to define GDM as being high or low (Table 1). This classification was also added as a covariate to the bivariate model.

We calculated summary estimates of accuracy measures using studies that reported on a threshold of 7.8 mmol/l. In order to evaluate accuracy measures over the whole range of possible thresholds, we estimated accuracy as a function of the 50-g glucose challenge test threshold values by including this value as a continuous covariate in the bivariate model. In order to avoid results being biased towards studies reporting on many different thresholds, we estimated the model in 250 stratified bootstrap samples, in which only one threshold value from each study was randomly selected. The final model was based on the average over all estimates from 250 bootstrap samples. The model parameters were used to produce sROC curves, where the increase in sensitivity and decrease in specificity reflect the shift in threshold value of the 50-g glucose challenge test in the model. Separate ROC curves reflect each type of study (studies with consecutive recruitment of patients versus studies including high-risk women, and low versus high OGTT threshold values). Statistical analyses were performed using SPSS 16.0 (SPSS, Chicago, IL, USA) and SAS 9.1.3 (SAS Institute Inc. Cary, NC, 2000-2004). Forest plots were made with Review Manager 5.0 (Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2008).

Results

Figure 1 summarises the process of literature identification and study selection. Our search resulted in 745 hits. We included 26 studies, comprising 13 564 women, of whom 1027 (7.5%) had GDM.[17-42]

Study characteristics

Table 2 summarises the characteristics of the studies included. Twenty-four studies (92.3%) were cohort studies, one was a randomised controlled trial and one was a cross-sectional study. All studies reported prospective recruitment. Sample sizes ranged from 42 women to 3836 women (median 378 women), with the prevalence of GDM varying from 3 to 33% (median 8%). Patient recruitment was reported as being consecutive in 21 studies (81%), whereas in four studies (15%) patients were screened based on the presence of risk factors. One study did not report inclusion criteria. Rates of GDM varied

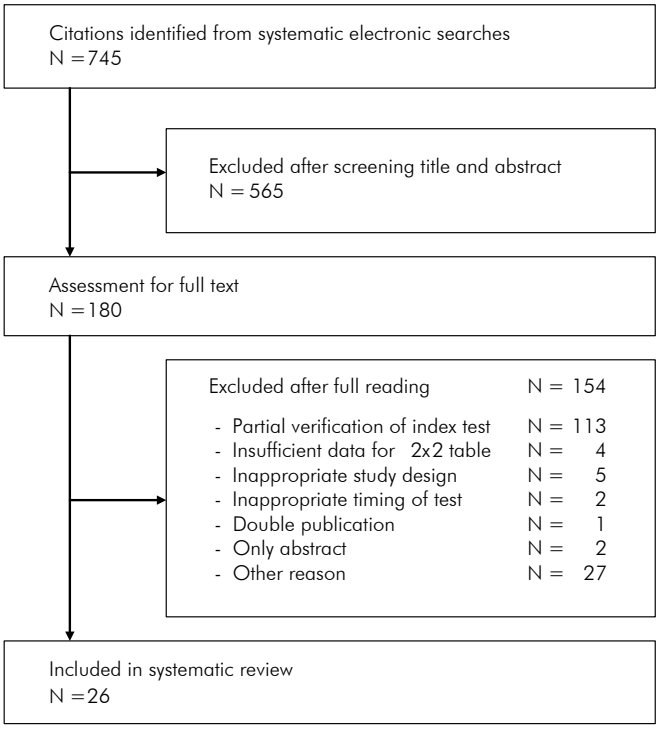


Figure 1. Literature identification and study selection.

from 3 to 33% in studies with consecutive recruitment, and from 11 to 17% in studies with inclusion based on the presence of risk factors.

Index test

All but two studies reported accuracy measures with the threshold of the 50-g glucose challenge test set at 7.8 mmol/l. Nineteen studies reported accuracy measures for multiple thresholds of the 50-g glucose challenge test. In the majority of the studies the 50-g glucose challenge test was performed between 24 and 28 weeks of gestation.

Reference test

In six studies the 75-g OGTT was used as a reference standard, whereas 20 studies used the 100-g OGTT as a reference standard. Ten studies were categorised as having a low threshold value of the OGTT and 16 studies were categorised as having a high threshold value of the OGTT to define GDM (Table 1 and 2). In the majority of the studies the OGTT was performed between 24 and 28 weeks of gestation.

Quality assessment

Figure 2 summarises the results of the methodological quality assessment. Inclusion criteria were reported in 25 studies (96.1%). Verification of the 50-g glucose challenge

Table 2. Key characteristics of included studies

First author, year	Country	Design	Inclusion	Exclusion	Gestational age (weeks)	Sample size (n)
Ayach, 2006.[17]	Brasil	Cohort	Consecutive	Preterm birth, fetal death	24 - 28	341
Bonomo, 1998.[18]	Italy	Cohort	Consecutive	None	24 - 28	704
Caliskan, 2004.[19]	Turkey	Cohort	Consecutive	None	24 - 28	422
Cetin, 1997.[20]	Turkey	Cohort	Consecutive	Medication, preterm birth, preeclampsia	24 - 28	274
Cocilovo, 1994.[21]	Italy	Cohort	Consecutive	None	24 - 30	249
Espinosa, 1999.[22]	Mexico	Cohort	Consecutive	None	24 - 30	445
Hidar, 2001.[23]	France	Cohort	Consecutive	Preterm birth, PPROM	24 - 28	95
Jirapinyo, 1993.[24]	Thailand	Cohort	Risk factors	None	8 - 38 65% 24-30	396
Keshavarz, 2006.[25]	Iran	Cohort	Consecutive	None	24 - 28	412
Lamar, 1999.[26]	USA	RCT	Consecutive	None	24 - 28	136
Maegawa, 2003.[27]	Japan	Cohort	Not reported	None	13 - 26	735*
Mathai, 1994.[28]	India	Cohort	Consecutive	Delivery elsewhere	26 - 30	232
Perea, 2002.[29]	Spain	Cohort	Consecutive	None	24 - 28	138
Perea, a, 2002.[29]	Spain	Cohort	Consecutive	None	24 - 28	642
Perucchini, 1999.[30]	Switzerland	Cohort	Consecutive	None	24 - 28	520
Puavilai, 1993.[31]	Thailand	Cohort	Risk factors	None	24 - 28	115
Ramirez, 2003.[32]	Mexico	Cohort	Consecutive	None	24 - 28	334
Rey, 2004.[33]	Canada	Cohort	Consecutive	None	24 - 28	188
Schwartz, 1994.[34]	USA	Cohort	Consecutive	None	7 - 40, mean 27.6	132
Sermer, 1994.[35]	Canada	Cohort	Consecutive	None	26 - 27	3836
Siribaddana, 1998.[36]	Sri Lanka	Cohort	Consecutive	None	24 - 28	721
Tam, 2000.[37]	China	Cohort	not reported	None	24 - 28	893
Thitadilok, 1995.[38]	Thailand	Cohort	Risk factors	None	24 - 28	304
Uncu, 1995.[39]	Turkey	Cohort	Consecutive	None	24 - 28	42
v. Leeuwen, 2007.[40]	The Netherlands	Cohort	Consecutive	None	24 - 28	1281
Weerakiet, 2006.[41]	Thailand	Cross sectional	Risk factors	Chronic disease (treatment)	21 - 28	359

GCT = 50 g glucose challenge test; GDM = gestational diabetes mellitus; OGTT = oral glucose tolerance test; * testing in the second trimester (women diagnosed in the first trimester were excluded); **For a threshold of the 50 g glucose challenge test of 7.2 mmol/L. Accuracy measures were not reported for a threshold of 7.8 mmol/L. *** For a threshold of the 50 g glucose challenge test of 7.0 mmol/L. Accuracy measures were not reported for a threshold of 7.8 mmol/L.

GCT verified (%)	GDM n (%)	OGTT 75/100 g	Threshold OGTT high / low	Sensitivity 7.8 mmol/L	Specificity 7.8 mmol/L	Capillary / venous and plasma / blood
100	13 (4)	100	Low	0.77	0.88	Plasma
100	50 (7)	100	High	0.91	0.73	Venous Plasma
99	14 (3)	100	High	0.93	0.76	Venous
100	17 (6)	100	High	0.65	0.88	Plasma
100	9	100	High	1.0	0.76	Capillary Blood
100	43 (10)	100	High	0.88	0.85	Venous
100	13 (14)	75	Low	0.69	0.87	Plasma
100	42 (11)	100	High	0.86	0.65	Plasma
90	26 (6)	100	Low	0.88	0.88	Unknown
100	5	100	High	0.8	0.82	Venous
100	8 (1*)	75	High	0.79	0.85	Unknown
100	11 (5)	100	Low	0.36	0.8	Plasma
100	13 (9)	100	High	1.00	0.77	Unknown
100	53 (8)	100	High	0.98	0.74	Unknown
93	53 (10)	100	Low	0.58	0.91	Venous Plasma
100	16 (14)	100	High	0.19	0.95**	Unknown
100	24 (7)	100	Low	0.88	0.64	Venous
78	21 (11)	75	High	0.79	0.97	Unknown
100	25 (19)	100	High	0.92	0.52	Venous plasma
100	265 (7)	100	High	0.77	0.82	Plasma
100	40 (6)	75	Low	0.63	0.84	Plasma
95	122 (13)	75	Low	0.73	0.68 ***	Unknown
100	23 (8)	100	High	0.91	0.77	Plasma
100	14 (33)	100	High	0.79	0.54	Plasma
98	47 (4)	75	Low	0.70	0.89	Venous plasma
100	60 (17)	100	Low	0.90	0.61	Plasma

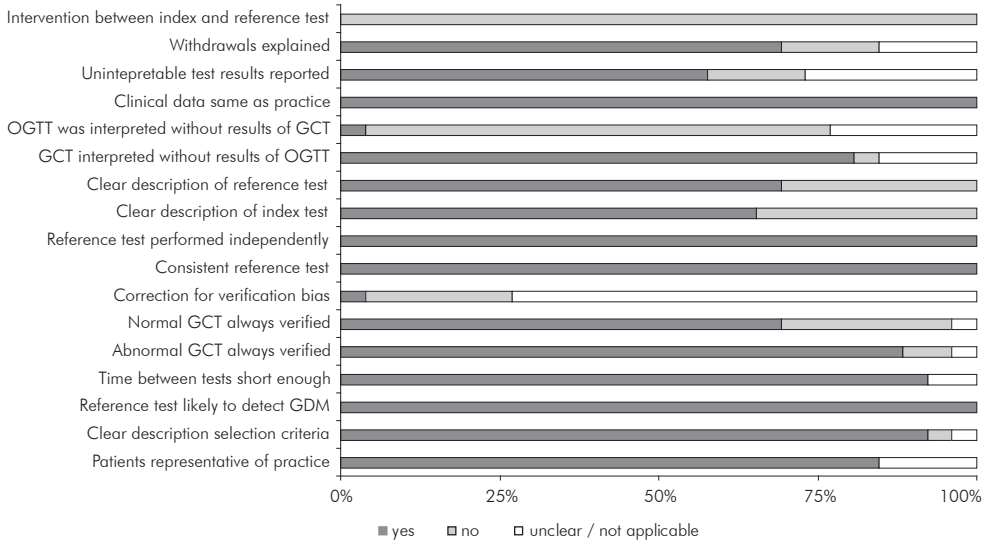


Figure 2. Methodological quality of the studies included in the systematic review of the 50-g glucose challenge test as a screening test for GDM.

test results was 100% in 20 studies (76.9%) and >90% in five studies (19%). Details on the administration of the index and the reference test were reported in 73.1% and 76.9% of the studies.

Data analysis

We could construct 125 2x2 tables. The sensitivity of the index test reported by the 26 individual studies ranged from 0 to 100%. The specificity ranged from 2 to 100%. The combined values of the sensitivity and specificity calculated from the 2x2 tables are plotted in Figure 3. The wide range of sensitivity and specificity was mainly a result of variation in threshold values of the index test used to define test positivity (range 4.0-16.0 mmol/l). Figure 4 shows sensitivity and specificity of the individual included studies for three commonly used thresholds. Studies reporting several thresholds appear multiple times in this chart.

First we estimated pooled sensitivity and specificity with a bivariate model in which we used data from the studies that reported accuracy measures of the index test for the threshold of 7.8 mmol/l. We evaluated the effect of covariates on sensitivity and specificity. The recruitment of patients was associated with specificity but not with sensitivity of the index test. The specificity of the index test was lower in studies that included women with risk factors compared with studies with consecutive recruitment ($P = 0.02$). There was no association between the type of reference test (75- or 100-g OGTT) and sensitivity or specificity. A higher threshold of the reference test for the diagnosis of GDM increased the sensitivity and decreased the specificity of the index test compared with a low threshold

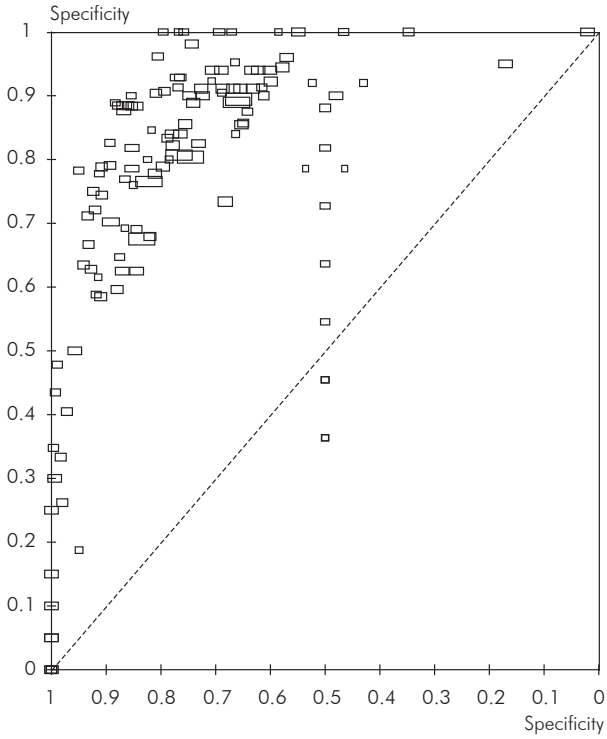


Figure 3. ROC plot of sensitivity and specificity for all studies included.

All studies are displayed irrespective of recruitment (consecutive or inclusion of women with risk factors), reference test (75- or 100-g OGTT), threshold value of the index test (high or low) and threshold value of the 50-g glucose challenge test. The width of the blocks is proportional to the inverse standard error of specificity. The height of the blocks is proportional to the inverse standard error of sensitivity.

(P = 0.01 and 0.01, respectively). For studies that included women with risk factors the pooled sensitivity was 0.76 (95% CI 0.67 - 0.84), with a pooled specificity of 0.76 (95% CI 0.60 - 0.87), and with a derived positive LR of 3.2 (95% CI 1.8 - 5.7) and negative LR of 0.32 (95% CI 0.21 - 0.47). For studies that included consecutively recruited women the pooled sensitivity was 0.76 (95% CI 0.60 - 0.87), with a pooled specificity of 0.85 (95% CI 0.81 - 0.88), and with a derived LR of 5.1 (95% CI 3.7 - 6.0) and negative LR of 0.28 (95% CI 0.16 - 0.51). With increasing the threshold of OGTT to diagnose GDM, the pooled sensitivity increased to 0.89 (95% CI 0.72 - 0.97), with a decreasing specificity of 0.77 (95% CI 0.63 - 0.86), for studies that recruited all pregnant women, and a specificity of 0.65 (95% CI 0.38 - 0.85) for studies that recruited women based on the presence of risk factors, with a derived positive LR of 3.9 (95% CI 2.3 - 6.6) and 2.5 (95% CI 1.2 - 5.6), and a derived negative LR of 0.14 (95% CI 0.05 - 0.43) and 0.17 (95% CI 0.05 - 0.55), respectively.

Next, we estimated accuracy measures for all reported threshold values by including the threshold value of the 50-g glucose challenge test as a covariate in the bivariate model. The effects of the covariates (patient recruitment, reference test, threshold value of the reference test) were the same as described above. The effects of threshold value of the 50-g glucose challenge test were statistically significant for sensitivity ($P < 0.0001$) as well as for specificity ($P < 0.0001$).

GCT threshold 7.5 mmol/L

Study	TP	FP	FN	TN	Sensitivity	Specificity
Bonomo 1998	11	15	3	13	0.79 [0.49-0.95]	0.46 [0.28-0.66]
Cocilovo 1994	9	79	0	161	1.00 [0.66-1.00]	0.67 [0.61-0.73]
Espinosa 1999	38	63	5	339	0.88 [0.75-0.96]	0.84 [0.80-0.88]
Mathai 1994	5	46	6	175	0.45 [0.17-0.77]	0.79 [0.73-0.84]
Perucchini 1999	32	56	21	411	0.60 [0.46-0.74]	0.88 [0.85-0.91]
Sermer 1994	117	860	28	2831	0.81 [0.73-0.87]	0.77 [0.75-0.78]
Uncu 1995	11	15	3	13	0.79 [0.49-0.95]	0.46 [0.28-0.66]

GCT threshold 7.8 mmol/L

Study	TP	FP	FN	TN	Sensitivity	Specificity
Ayach 2006	10	44	3	284	0.77 [0.46-0.95]	0.87 [0.82-0.90]
Bonomo 1998	47	203	3	451	0.94 [0.83-0.99]	0.69 [0.65-0.72]
Caliskan 2004	13	96	1	312	0.93 [0.66-1.00]	0.76 [0.72-0.81]
Cetin 1997	11	32	6	225	0.65 [0.38-0.86]	0.88 [0.83-0.91]
Cocilovo 1994	9	58	0	182	1.00 [0.66-1.00]	0.76 [0.70-0.81]
Espinosa 1999	38	59	5	343	0.88 [0.75-0.96]	0.85 [0.81-0.89]
Hidar 2001	9	11	4	71	0.69 [0.39-0.91]	0.87 [0.77-0.93]
Jirapinyo 1993	36	124	6	230	0.86 [0.71-0.95]	0.65 [0.60-0.70]
Keshavarz 2006	23	47	3	339	0.88 [0.70-0.98]	0.88 [0.84-0.91]
Lamar 1999	4	23	1	108	0.80 [0.28-0.99]	0.82 [0.75-0.89]
Maegawa 2003	18	107	4	620	0.82 [0.60-0.95]	0.85 [0.82-0.88]
Mathai 1994	4	44	7	177	0.36 [0.11-0.69]	0.80 [0.74-0.85]
Perea 2002	52	151	1	438	0.98 [0.90-1.00]	0.74 [0.71-0.78]
Perea 2002a	13	29	0	96	1.00 [0.75-1.00]	0.77 [0.68-0.84]
Perucchini 1999	31	42	22	425	0.58 [0.44-0.72]	0.91 [0.88-0.93]
Puavilai 1993	21	111	3	199	0.88 [0.68-0.97]	0.64 [0.59-0.70]
Ramirez 2003	3	5	13	94	0.19 [0.04-0.46]	0.95 [0.89-0.98]
Rey 2004	20	56	1	111	0.95 [0.76-1.00]	0.66 [0.59-0.74]
Schwartz 1994	23	51	2	56	0.92 [0.74-0.99]	0.52 [0.42-0.62]
Sermer 1994	111	657	34	3034	0.77 [0.69-0.83]	0.82 [0.81-0.83]
Siribaddana 1998	25	106	15	575	0.63 [0.46-0.77]	0.84 [0.81-0.87]
Tam 2000	80	249	29	535	0.73 [0.64-0.81]	0.68 [0.65-0.71]
Thitadilok 1995	21	65	2	216	0.91 [0.72-0.99]	0.77 [0.71-0.82]
Uncu 1995	11	13	3	15	0.79 [0.49-0.95]	0.54 [0.34-0.72]
van Leeuwen 2007	33	134	14	1100	0.70 [0.55-0.83]	0.89 [0.87-0.91]
Weerakiet 2006	54	116	6	183	0.90 [0.79-0.96]	0.61 [0.55-0.67]

GCT threshold 8.0 mmol/L

Study	TP	FP	FN	TN	Sensitivity	Specificity
Bonomo 1998	77	150	13	464	0.86 [0.77-0.92]	0.76 [0.72-0.79]
Cocilovo 1994	9	49	0	191	1.00 [0.66-1.00]	0.80 [0.74-0.84]
Espinosa 1999	41	36	11	357	0.79 [0.65-0.89]	0.91 [0.88-0.94]
Siribaddana 1998	25	88	15	593	0.63 [0.46-0.77]	0.87 [0.84-0.90]

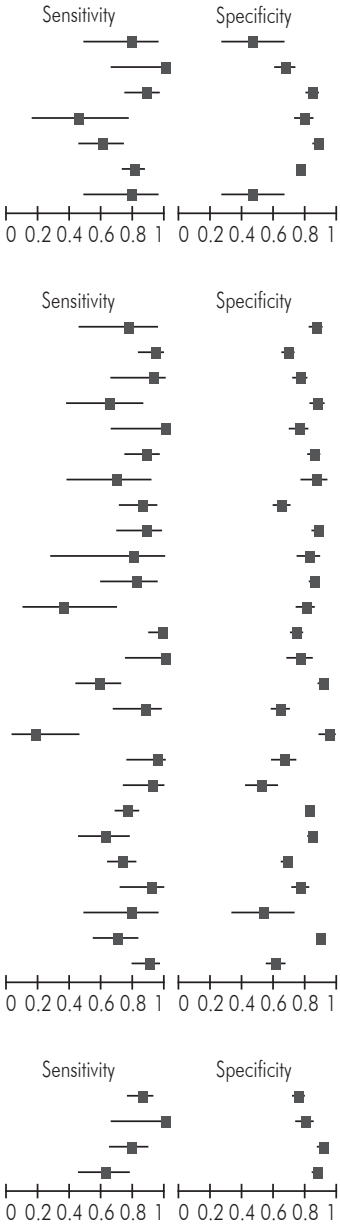


Figure 4. Forest plots for all studies, shown for three thresholds that are frequently applied in current clinical practice (7.5 mmol/L, 7.8 mmol/L and 8.0 mmol/L). Squares represent sensitivity and specificity. The black line is the confidence interval. Some studies are shown multiple times for several thresholds.

The pooled sensitivity and specificity for three threshold values are presented in Table 3. For an index test threshold of 7.8 mmol/l, the pooled sensitivity in studies that recruited only women with risk factors was 0.74 (95% CI 0.62 - 0.87), with a pooled specificity of 0.77 (95% CI 0.66 - 0.89), and with a derived positive LR of 3.2 (95% CI 2.0 - 5.2) and negative LR 0.34 (95% CI 0.22 - 0.53). The pooled sensitivity of studies including all pregnant women was 0.74 (95% CI 0.62 - 0.87), with a pooled specificity of 0.85 (95% CI 0.80 - 0.91), with a derived positive LR of 4.9 (95% CI 3.5 - 7.0) and negative LR of 0.31 (95% CI 0.20 - 0.47). Increasing the threshold of the reference test for the diagnosis of GDM increased sensitivity to 0.83 (95% CI 0.75 - 0.91), and decreased specificity to 0.81 (95% CI 0.75 - 0.87) for studies that recruited all pregnant women (derived positive LR 4.4 (95% CI 3.2 - 6.0); negative LR 0.21 (95% CI 0.14 - 0.32)), and 0.72 (95% CI 0.60 - 0.84) for studies that recruited women based on the presence of risk factors (derived positive LR 3.0 (95% CI 2.0 - 4.5); negative LR 0.24 (95% CI 0.15 - 0.37)). Summary ROC curves that reflect the results for all possible thresholds of the 50-g glucose challenge test are displayed in Figure 5.

Table 3. Accuracy measures for three thresholds of the 50 g glucose challenge test estimated with a bivariate regression model.

	Recruitment of all women		Recruitment of women with risk factors	
	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
OGTT with low threshold for disease				
7.5 mmol/L	0.78 (0.67-0.89)	0.81 (0.74-0.87)	0.78 (0.67-0.89)	0.72 (0.58-0.85)
7.8 mmol/L	0.74 (0.62-0.87)	0.85 (0.80-0.91)	0.74 (0.62-0.87)	0.77 (0.66-0.89)
8.0 mmol/L	0.72 (0.59-0.85)	0.88 (0.83-0.93)	0.72 (0.59-0.85)	0.81 (0.71-0.91)
OGTT with high threshold for disease				
7.5 mmol/L	0.85 (0.78-0.93)	0.76 (0.69-0.83)	0.85 (0.78-0.93)	0.65 (0.51-0.79)
7.8 mmol/L	0.83 (0.75-0.91)	0.81 (0.75-0.87)	0.83 (0.75-0.91)	0.72 (0.60-0.84)
8.0 mmol/L	0.81 (0.72-0.90)	0.84 (0.79-0.89)	0.81 (0.72-0.90)	0.76 (0.65-0.86)

The threshold of the 50 g glucose challenge test was added as a covariate. Recruitment of women (recruitment of all women versus recruitment of women with risk factors for gestational diabetes mellitus (GDM)) was added a covariate to the model. Threshold of the OGTT was also added as a covariate to the model.

Discussion

We systematically reviewed the literature on the accuracy of the 50-g glucose challenge test for the diagnosis of GDM. We found that the pooled estimate of sensitivity for a threshold value of 7.8 mmol/l ranged between 0.74 (95% CI 0.62 - 0.87) and 0.83 (95% CI 0.75 - 0.91), depending on the threshold for diagnosis of GDM, with a specificity ranging between 0.72 (95% CI 0.60 - 0.84) and 0.85 (95% CI 0.80 - 0.91).
Depending on the application of the test (screening or alternative diagnostic) and the consequences of false-positive and false-negative test results, certain combinations of

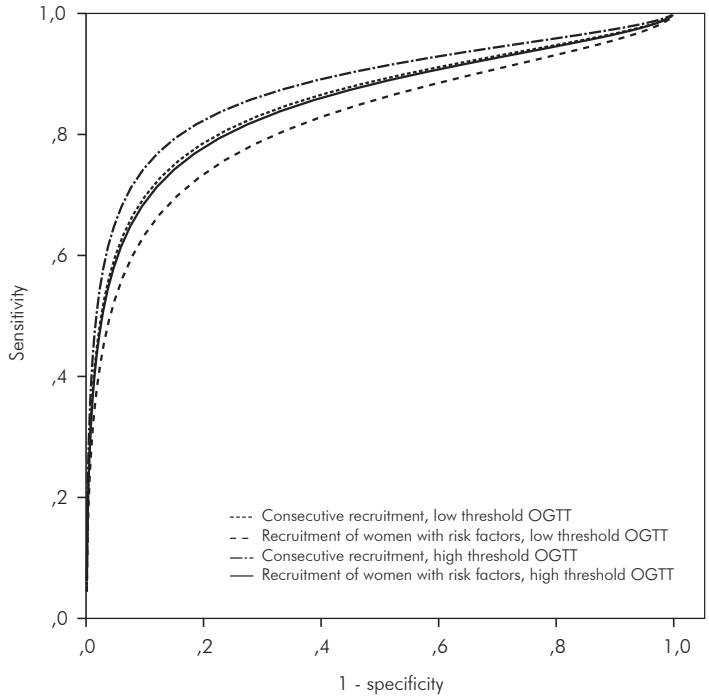


Figure 5. sROC plots of sensitivity and specificity for various subgroups (depending on recruitment and threshold of the reference test), based on the estimates of the bivariate model.

accuracy values are preferred. These values depend on whether it is more harmful to classify women as false-positive or false-negative, taking all possible consequences of such results into account. In the case of GDM, regarding the nature and consequences of the disease, one should aim for an adequate detection rate, albeit not at the cost of an unacceptable false-positive rate. If the 50-g glucose challenge test is used as a screening test, a higher sensitivity rate than 74% would probably be warranted to accept a false-positive rate of 83%. Moreover, if one considers using the 50-g glucose challenge test as a diagnostic test for GDM, higher detection rates are required. As the prevalence of GDM in the general population is relatively low, a clinically useful test would thus have to have a high positive LR (>10) and a low negative LR (<0.10). Regarding the derived positive and negative LRs in the present study, the accuracy of the 50-g glucose test for gestational diabetes mellitus is modest. For example, if the incidence of GDM is 3%, and the positive LR of the 50-g glucose challenge test is 4.5, the post-test probability for an abnormal result on the 50-g glucose challenge test would be 12%, which is still low. If the 50-g glucose challenge test is combined with other screening methods, such as the presence of risk factors for GDM (e.g. GDM in a previous pregnancy, obesity), LRs of the risk factors are multiplied with the LR of the 50-g glucose challenge test, and thus the post-test probability might be improved.

Our findings imply that if the 50-g glucose challenge test is performed in a cohort of 1000 unselected women, with an assumed prevalence of 4%, and with a threshold of 7.8 mmol/l, 30 women would have a true-positive test result and 144 would have a false-positive test result (positive predictive value of 17%). Ten women would have a false-negative test result and 816 women would have a true-negative test result (negative predictive value of 99%). In many potentially relevant studies dealing with the 50-g glucose challenge test in pregnant women, the OGTT was only performed if the 50-g glucose challenge test was considered to be abnormal. This design characteristic, known as partial verification, is encountered in many studies on diagnostic accuracy: to minimise the burden of possibly redundant additional testing in women with a negative screening test result, only abnormal screening test results are verified by the reference test. To avoid this partial verification bias, only studies that performed both a 50-g glucose challenge test and an OGTT independent of the result of the gram glucose challenge test in more than 75% of the women were included in this systematic review.

A limitation of the present study is that details on study and sample characteristics were not reported equally well in the individual studies. Because of this incomplete reporting, we were not able to evaluate every quality item in every study. Inclusion of methodologically poor studies may have affected our estimates of (diagnostic accuracy).[42] With the bivariate model that we used to calculate summary estimates of sensitivity and specificity, the potential influence of clinical and study characteristics (covariates) on the mean sensitivity and specificity can be evaluated. Because of the limited details that were reported, clinical variables of interest (for example age, BMI and time between the last meal and the index test), the number of covariates included in the bivariate model was limited. To evaluate the true effect of clinical variables and different threshold values, individual patient data meta-analysis is needed.

Another limitation of our study was the lack of a uniform reference test for GDM. The glucose load of the OGTT is either 75 or 100 g. Both tests are used in clinical practice. For the 75-g as well as for the 100-g OGTT, various criteria for an abnormal test result exist. Direct comparison between all studies included in our systematic review was therefore complicated. With the bivariate model we were able to account for the variation in summary estimates caused by the difference in cutoff values used for the index test. We accounted for variation in summary estimates caused by the various criteria to define an abnormal OGTT as well, by categorising the threshold for an abnormal OGTT as being high or low. Adding this variable to the bivariate model increased the fit of the model. We do not know, however, to what extent the arbitrary categorisation of the OGTT thresholds is justified.
In pregnancy placental hormones cause maternal insulin sensitivity to decrease, and as a consequence postprandial glucose levels increase. Combs *et al*.[43] showed that rising

postprandial glucose values were associated with fetal macrosomia, a common feature in pregnancies complicated by GDM. A glucose loading test like the 50-g glucose challenge test in theory seems an adequate method to mimic postprandial glucose levels, and therefore to measure the degree of glucose (in) tolerance in pregnancy.

A health technology report concerning various screening strategies for GDM stated that the cost-effectiveness of a of number of studies find that screening with the 50-g glucose challenge test, and then testing screen-positives with the OGTT, was less costly than going straight to universal OGTT. However, a high-quality cost-effectiveness analysis developed by the UK's National Institute for Health and Clinical Excellence (NICE) guideline development group, that compared costs and effects of multiple screening strategies including the 50-g glucose challenge test, found that two other screening strategies dominated: selection by American Diabetes Association (ADA) risk criteria, followed by the 75-g OGTT; and selection by high-risk ethnicity, followed by the 75-g OGTT.[44] In view of these findings, and as an extension to the results of the cost-effectiveness analysis of the NICE guideline development group, it would be interesting to consider the cost effectiveness of a strategy that consists of selection based on various risk factors, followed by screening with a 50-g glucose challenge test, followed by an OGTT in the case of an abnormal test result of the 50-g glucose challenge test, and to compare this in a randomised controlled trial with other screening strategies.

Conclusion

As GDM is often asymptomatic, screening is necessary to identify women with GDM. High sensitivity is often warranted in screening tests, as a false-negative test result (in which disease remains undiscovered) is considered to be more harmful than a false-positive test result (in which a reference test is unnecessarily performed). Although higher detection rates would be preferable, the detection rate of the 50-g glucose challenge test of 74% might be acceptable if used as a screening test for a condition such as GDM. On the other hand one could consider a one-step method, using the OGTT for screening for example in a selected population (risk factors). This could be a lesser burden for women and more cost-effective than a two-step method in which a glucose-loading test might be performed twice. To use the 50-g glucose challenge test as a definite diagnostic test for GDM (replacement of the OGTT), higher accuracy measures are warranted.

References

1. Jensen DM, Damm P, Sørensen B, Mølsted-Pedersen L, Westergaard JG, Korsholm L, et al. Proposed diagnostic thresholds for gestational diabetes mellitus according to a 75-g oral glucose tolerance test. Maternal and perinatal outcomes in 3260 Danish women. Diabet Med 2003; 20:51-7.
2. Esakoff TF, Cheng YW, Caughey AB. Screening for gestational diabetes: different cut-offs for different ethnicities? Am J Obstet Gynecol 2005; 193:1040-4.
3. Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS; Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) Trial Group. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med 2005; 352:2477-86.
4. Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR, et al.; HAPO Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008; 358:1991-2002.
5. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. Diabetes Care 2002; 25:1862-8.
6. Plagemann AJ. A matter of insulin: developmental programming of body weight regulation. Matern Fetal Neonatal Med 2008; 21:143-8.
7. Tam WH, Ma RC, Yang X, Li AM, Ko GT, Kong AP, et al. Glucose intolerance and cardiometabolic risk in adolescents exposed to maternal gestational diabetes: a 15-year follow-up study. Diabetes Care 2010; 33:1382-4.
8. Langer O, Yogev Y, Most O, Xenakis EM. Gestational diabetes: the consequences of not treating. Am J Obstet Gynecol 2005; 192:989-97.
9. Landon MB, Spong CY, Thom E, Carpenter MW, Ramin SM, Casey B, et al. A multicenter, randomized trial of treatment for mild gestational diabetes. N Engl J Med 2009; 361:1339-48.
10. Gabbe SG, Gregory RP, Power ML, Williams SB, Schulkin J. Management of diabetes mellitus by obstetrician-gynecologists. Obstet Gynecol 2004; 103:1229-34.
11. Leeflang MM, Scholten RJ, Rutjes AW, Reitsma JB, Bossuyt PM. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. J Clin Epidemiol 2006; 59:234-40.
12. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282:1061-6.
13. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med 2004; 140:189-202.
14. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003; 3:25.
15. Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med Res Methodol 2005; 5:19.
16. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005; 10:982-90.
17. Ayach W, Costa RA, Calderon Ide M, Rudge MV. Comparison between 100-g glucose tolerance test and two other screening tests for gestational diabetes: combined fasting glucose with risk factors and 50-g glucose tolerance test. Sao Paulo Med J 2006; 124:4-9.
18. Bonomo M, Gandini ML, Mastropasqua A, Begher C, Valentini U, Faden D, et al. Which cutoff level should be used in screening for glucose intolerance in pregnancy? Definition of Screening Methods for Gestational Diabetes Study Group of the Lombardy Section of the Italian Society of Diabetology. Am J Obstet Gynecol 1998; 179:179-85.

19. Caliskan E, Kayikcioglu F, Ozturk N, Koc S, Haberal A. A population-based risk factor scoring will decrease unnecessary testing for the diagnosis of gestational diabetes mellitus. Acta Obstet Gynecol Scand 2004; 83:524-30.
20. Cetin M, Cetin A. Time-dependent gestational diabetes screening values. Int J Gynaecol Obstet 1997; 56:257-61.
21. Cocilovo G, Leonardi MC, Vesce F, Mollica G. Cut-off values of laboratory measurements of capillary blood glucose in a glucose challenge test. Evaluation of 250 cases. J Obstet Gynaecol 1994; 14:76-8.
22. Espinosa de los Monteros A, Parra A, Carino N, Ramirez A. The reproducibility of the 50-g, 1-hour glucose screen for diabetes in pregnancy. Obstet Gynecol 1993; 82:515-8.
23. Hidar S, Chaıeb A, Baccouche S, Laradi S, Fkih M, Milled A, et al. Post-prandial plasma glucose test as screening tool for gestational diabetes: a prospective randomized trial. J Gynecol Obstet Biol Reprod (Paris) 2001; 30:344-7.
24. Jirapinyo M, Puavilai G, Chanprasertyotin S, Tangtrakul S. Predictive value of 1 hour 50 g oral glucose load screening test for gestational diabetes mellitus compared to 3 hour oral glucose tolerance test in high risk pregnant women. Asia Oceania J Obstet Gynaecol 1993; 19:7-12.
25. Keshavarz M, Cheung NW, Babaee GR. The value of the screening test for gestational diabetes in pregnant Iranian women. Diabetes Res Clin Pract 2006; 73:98-9.
26. Lamar ME, Kuehl TJ, Cooney AT, Gayle LJ, Holleman S, Allen SR. Jelly beans as an alternative to a fifty-gram glucose beverage for gestational diabetes screening. Am J Obstet Gynecol 1999; 181:1154-7.
27. Maegawa Y, Sugiyama T, Kusaka H, Mitao M, Toyoda N. Screening tests for gestational diabetes in Japan in the 1st and 2nd trimester of pregnancy. Diabetes Res Clin Pract 2003; 62:47-53.
28. Mathai M, Thomas TJ, Kuruvila S, Jairaj P. Random plasma glucose and the glucose challenge test in pregnancy. Natl Med J India 1994; 7:160-2.
29. Perea-Carrasco R, Perez-Coronel R, Albusac-Aguilar R, Lombardo- Grifol M, Bassas-Baena de Leon E, Romero-Diaz C. A simple index for detection of gestational diabetes mellitus. J R Soc Med 2002; 95:435-9.
30. Perucchini D, Fischer U, Spinas GA, Huch R, Huch A, Lehmann R. Using fasting plasma glucose concentrations to screen for gestational diabetes mellitus: prospective population based study. BMJ 1999; 319:812-5.
31. Puavilai G, Chanprasertyotin S, Jirapinyo M. An evaluation of glycosylated hemoglobin measurement by a colorimetric method as a screening test for gestational diabetes mellitus. J Med Assoc Thai 1993; 76:549-53.
32. Ramı□rez-Torres MA, Rodrıguez-Pezino J, Zambrana-Castaneda M, Lira-Plascencia J, Parra A. Gestational diabetes mellitus and glucose intolerance among Mexican pregnant adolescents. J Pediatr Endocrinol Metab 2003; 16:401-5.
33. Rey E, Hudon L, Michon N, Boucher P, Ethier J, Saint-Louis P. Fasting plasma glucose versus glucose challenge test: screening for gestational diabetes and cost effectiveness. Clin Biochem 2004; 37:780-4.
34. Schwartz JG, Phillips WT, Blumhardt MR, Langer O. Use of a more physiologic oral glucose solution during screening for gestational diabetes mellitus. Am J Obstet Gynecol 1994; 171:685-91.
35. Sermer M, Naylor CD, Gare DJ, Kenshole AB, Ritchie JW, Farine D, et al. Impact of time since last meal on the gestational glucose challenge test. The Toronto Tri-Hospital Gestational Diabetes Project. Am J Obstet Gynecol 1994; 171:607-16.
36. Siribaddana SH, Deshabandu R, Rajapakse D, Silva K, Fernando DJ. The prevalence of gestational diabetes in a Sri Lankan antenatal clinic. Ceylon Med J 1998; 43:88-91.
37. 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; 23:1432.

38. Thitadilok W, Techatraisak K. The screening for gestational diabetes employing 50 g glucose test. J Med Assoc Thai 1995; 78:526-31.
39. Uncu G, Ozan H, Cengiz C. The comparison of 50 grams glucose challenge test, HbA1c and fructosamine levels in diagnosis of gestational diabetes mellitus. Clin Exp Obstet Gynecol 1995; 22:230-4.
40. van Leeuwen M, Zweers EJ, Opmeer BC, van Ballegooie E, ter Brugge HG, de Valk HW, et al. Comparison of accuracy measures of two screening tests for gestational diabetes mellitus. Diabetes Care 2007; 30: 2779-84.
41. Weerakiet S, Lertnarkorn K, Panburana P, Pitakitronakorn S, Vesathada K, Wansumrith S. Can adiponectin predict gestational diabetes? Gynecol Endocrinol 2006; 22:362-8.
42. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006; 174:469-6.
43. Combs CA, Gunderson E, Kitzmiller JL, Gavin LA, Main EK. Relationship of fetal macrosomia to maternal postprandial glucose control during pregnancy. Diabetes Care 1992; 15:1251-7.
44. Waugh N, Royle P, Clar C, Henderson R, Cummins E, Hadden D, Lindsay R, Pearson D. Screening for hyperglycemia in pregnancy: a rapid update for the National Screening Committee. Health Technol Assess. 2010; 14(45):1-183.

Chapter 6

External validation of a clinical scoring system for the risk of gestational diabetes mellitus

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

Diabetes Research and Clinical Practice. 2009; 85(1):96-101

Abstract

AIM A prediction rule for gestational diabetes mellitus (GDM) could be helpful in early detection and increased efficiency of screening. A prediction rule by means of a clinical scoring system is available, but has never been validated externally. The aim of this study was to validate the scoring system.

METHODS We used data from a prospective cohort study. Women were assigned a score based on age, Body Mass Index (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). We compared the efficiency of a screening strategy derived from the scoring system with conventional screening.

RESULTS We studied 1266 women. Forty-seven women had GDM (3.7%). The scoring system discriminated moderately (area under the curve = 0.64 (95% CI 0.56 - 0.72)). Calibration was limited ($\chi 2 = 8.89$, $p = 0.06$). The screening strategy derived from the scoring system reduced the number of women needed to be screened with 25% for a comparable detection rate to universal screening.

CONCLUSION Despite moderate discriminative capacity and calibration of the scoring system, the screening strategy based on the scoring system appears clinically useful. There is need for better prediction models for GDM.

Introduction

In pregnancies complicated by gestational diabetes mellitus (GDM), there is an increased rate of fetal as well as maternal complications during pregnancy and delivery. Early diagnosis and subsequent treatment of GDM could prevent these complications and therefore improve pregnancy outcome.[1] GDM often is an asymptomatic condition. The optimal strategy to identify women with GDM is unknown. Some expert groups advocate the use of clinical factors to identify women at risk for GDM.[2] If clinical factors can be used to estimate the probability of GDM accurately, discrimination between women at high risk and women at low risk can be made early in pregnancy or even before conception. Additional screening procedures could then be limited to women at increased risk for GDM. This would reduce the burden of screening on low risk women, whereas women at high risk could be monitored closely and treated in a timely fashion if necessary.

The probability that disease will occur can be estimated with a prognostic model or clinical scoring system. In 1997, Naylor *et al*. published a scoring system for the risk classification of GDM that was based on multivariable logistic regression analysis of clinical characteristics.[3] Based on the variables age, Body Mass Index (BMI) and ethnicity a score was calculated for all women in the study. According to this score, women were classified as being at low, intermediate or high risk for GDM, and a screening strategy was developed based on this classification. Internal validation of the scoring system showed that the score successfully differentiated between women at high risk and women at low risk for GDM. Naylor *et al*. used the 100-g oral glucose tolerance test (OGTT) to establish the diagnosis of GDM. In current clinical practice however, GDM is often diagnosed using the 75-g OGTT instead of the 100-g OGTT. At present it is unknown if the scoring system by Naylor *et al*. is valid when using the 75 g-OGTT to diagnose GDM. The aim of this study was to validate the scoring system for GDM in an external population using the 75-g OGTT as a test for diagnosis of GDM. Moreover, we evaluated the efficiency of a screening strategy that was based on the scoring system.

Material and Methods

We used data from a previously published prospective cohort study, which compared the performance of two screening tests for GDM.[4] In this study, consecutive women with a singleton pregnancy who reported for prenatal care before 24 weeks of gestation in two hospitals in the Netherlands (the Isala Clinics in Zwolle and the University Medical Centre in Utrecht) were invited to participate. A random glucose measurement was performed in all women at intake (around 12 weeks of gestation), to detect women with undiagnosed type-1 or type-2 diabetes before pregnancy. Women with a diagnosis of pre-existing

diabetes mellitus were excluded. Medical history and characteristics of women who gave informed consent were recorded at intake. At 24 to 28 weeks of gestation all women underwent a 50-g glucose challenge as well as random glucose measurement to screen for GDM. The predefined cutoff value for an abnormal test result for the 50-g glucose challenge test was a 1-h plasma glucose value of 7.8 mmol/l. The random glucose test was considered as abnormal if the plasma glucose value ≥ 6.8 mmol/l. Women with an abnormal result on either or both screening tests underwent the 75-g oral glucose tolerance test (OGTT, reference test) within one week of the screening tests to confirm or rule out GDM. GDM was diagnosed if the fasting value of the OGTT was > 7.0 mmol/L or the two hour value of the OGTT was ≥ 7.8 mmol/L, according to the criteria of the World Health Organization (WHO).[5]

If all women with two negative screening tests would automatically be considered GDM negative, incidental cases of GDM would remain undetected, generating verification bias. To correct for verification bias the OGTT was performed in a subset of women with two negative screening test results, to estimate the fraction of diseased women that remained undetected by the screening tests (false-negative fraction). Subsequently, missing OGTT values due to the selective verification were imputed with a multiple imputation procedure. Details of this procedure have been reported in a previously published study.[4] At the same time, other incidental missing data on continuous variables were also imputed using this multiple imputation procedure.

Table 1 shows the clinical scoring system developed by Naylor *et al*. The scoring system is based on converted odds ratios derived by multivariable logistic regression analysis and includes three clinical variables: age, BMI and ethnicity. Based on these variables, women are assigned a clinical risk score, with a maximum score of 10 points (Table 1). Subsequently, women are classified into three categories. Women with a clinical risk score of 0 or 1 are categorized as low risk, women with a clinical risk score of 2 or 3 are categorized as intermediate risk and women with score higher than 3 are categorized as high risk for GDM. The screening strategy based on this clinical scoring system is as follows: Low risk women are not screened. Intermediate risk women are screened with the 50-g glucose challenge test with a threshold of 7.8 mmol/L. High risk women are also screened with the 50-g glucose challenge test, however a lower threshold is set for test positivity (7.1 mmol/L).

We used the clinical scoring system to calculate individual clinical risk scores for all women in our cohort and categorized women as being low, intermediate or high risk according to the definition by Naylor *et al*. We compared the overall prevalence of GDM between our sample and the sample of Naylor *et al*. and evaluated the distribution of women over the clinical risk scores and risk categories. We assessed the validity of the

Table 1. Scoring system according to Naylor *et al*. representing independent clinical risk factors for gestational glucose intolerance with odds ratios and corresponding scores derived from a multiple logistic regression model.

Risk Factor	Odds Ratio (95% CI)	P-value	Score according to Naylor *et al*.
Age (reference category ≤ 30 yr)			0
31 – 34 yr	1.0 (0.7 - 1.5)	0.95	1
≥ 35 yr	1.6 (1.1 - 2.5)	0.02	2
BMI (reference category ≤ 22.0)			0
22.1 – 25.0	1.8 (1.1 - 2.7)	0.01	2
≥ 25.1	3.2 (2.1 - 4.8)	<0.001	3
Ethnicity (reference category Caucasian)			0
Black	0.7 (0.3 - 1.7)	0.44	0
Asian	4.8 (3.0 - 7.6)	<0.001	5
Other	1.6 (0.7 - 3.5)	0.24	2

clinical scoring system by means of discrimination and calibration. Finally we evaluated the accuracy of the screening strategy based on the scoring system. Discrimination was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC). In the ROC plot the false-negative fraction was plotted against the true-positive fraction for all possible threshold values of the clinical risk score. The area under the ROC curve evaluated the ability of the clinical scoring system to distinguish women with GDM from women without GDM. The larger the AUC of the ROC curve, the better the discriminative capacity of the scoring system. An AUC of 0.5 indicates that the scoring system does no better than chance in estimating the outcome, whereas an AUC of 1.0 reflects perfect discriminative capacity. Calibration was evaluated with the χ^2 goodness of fit test to assess the level of correspondence between predicted probabilities and the observed percentage of women with GDM per clinical risk score. If the observed percentages of GDM are close to the predicted probabilities, the scoring system is considered to be well calibrated. We used the observed percentages of women with GDM in the original patient sample of Naylor *et al*. as the predicted probabilities of GDM. The goodness of fit test therefore reflected whether the prevalence of GDM across risk scores in our sample was statistically different from the prevalence reported by Naylor *et al*.

We evaluated the accuracy of the proposed screening strategy in terms of number of women needed to screen to establish a diagnosis of GDM, detection rate and false-positive rate, and compared these figures to those of universal screening with the 50-g glucose challenge test in our own sample using McNemar's test to test for agreement in screening accuracy. We also compared the accuracy measures to the figures that Naylor *et al*. found in their sample. In their original paper, Naylor *et al*. described two samples of women. The first sample was used to develop the scoring system and the subsequent screening strategies. We will refer to this group of women as Naylor's derivation sample.

The other sample was used to internally validate the scoring system and the developed screening strategy. We will refer to this sample as Naylor's validation sample. We used Naylor's derivation sample for comparison of our findings to the findings of Naylor *et al.* and only if data on Naylor's derivation sample were not available, we used Naylor's validation sample for comparison of results instead. Statistical analyses were performed using SPSS version 14.0.2 (SPSS, Chicago, IL) and SAS 9.1.3. No approval of the institutional research board was required for this validation study.

Results

There were 1301 women included in the original cohort study. Information concerning ethnicity was unavailable for 35 women and therefore these women were excluded from the analysis. Of 1266 women eligible for analysis, all women underwent the random glucose test and 1246 women underwent the 50-g glucose test (98.4%). 184 women had at least one abnormal screening test of whom 146 (80%) underwent an OGTT. 38 women did not agree to undergo an OGTT despite at least one abnormal result of the screening tests. To estimate the fraction of false negative screening test results, women with negative screening test results were asked at random to undergo the OGTT. 176 women consented and underwent an OGTT. The false-negative fraction was 7.9%. In total, the OGTT was performed in 322 women (25.4%), of whom 46 women had an abnormal OGTT result. After the multiple imputation procedure to correct for verification bias the number of women diagnosed with GDM was supposed to be 47, indicating an incidence of GDM in our sample of 3.7%. Next to the missing OGTT results, 7.3% of the

Table 2. Baseline characteristics of our sample with regard to the variables of the clinical scoring system.

	GDM present N = 47	GDM not present N = 1219	Total N = 1266
Age (years)			
≤ 30	26 (55.3)	588 (48.2)	614 (48.5)
31 - 34	7 (14.9)	342 (28.1)	349 (27.6)
≥ 35	14 (29.8)	289 (23.7)	303 (23.9)
BMI (kg/m2)			
≤ 22.0	8 (17.0)	433 (35.5)	441 (34.8)
22.1 - 25.0	9 (19.2)	398 (32.7)	407 (32.2)
≥ 25.1	30 (63.8)	388 (31.8)	418 (33.0)
Ethnicity			
Caucasian	38 (80.9)	1094 (89.8)	1132 (89.4)
Black	3 (6.3)	28 (2.3)	31 (2.5)
Asian	0 (0.0)	5 (0.4)	5 (0.4)
Other	6 (12.8)	92 (7.5)	98 (7.7)

Data are n (%).

data on age and BMI was imputed using the multiple imputation procedure, as well as the missing results of the 50-g glucose challenge test (1.6%). The baseline characteristics of our sample concerning the variables of the clinical scoring system are displayed in Table 2.

The overall prevalence of GDM in our sample of 3.7 % was not significantly different from the prevalence of GDM in Naylor's derivation group (2.8%, p= 0.18). Distribution of women over the risk scores and the observed percentages of GDM are shown in Figure 1. The distribution of women in our sample across the risk scores was significantly different from the distribution of women in Naylor's validation sample ($p < 0.001$). In Naylor's validation sample more women were in the low as well as in the high risk scores compared to our sample. The distribution of women across the predefined risk categories with their observed percentages of GDM are shown in Table 3. The distribution of women across the three risk categories in our population was significantly different from the distribution in the validation sample of Naylor *et al*. ($p < 0.001$).

The ROC curve of the clinical scoring system in our population had an AUC of 0.64 (95% CI 0.56 - 0.72), indicating a moderate discriminative capacity. The AUC in our sample was not significantly different from the AUC in the initial study of Naylor *et al*. (derivation group), which was 0.68. Since there were no results reported on the prevalence of GDM for the various clinical risk scores in Naylor's derivation sample, we compared the prevalence of GDM in our sample with the prevalence of GDM in the internal validation sample of Naylor *et al.* for the various clinical risk scores to assess the calibration of the model. Figure 1 shows the correspondence between the calculated risk score and the observed probability of GDM for our sample and for Naylor's validation sample. The

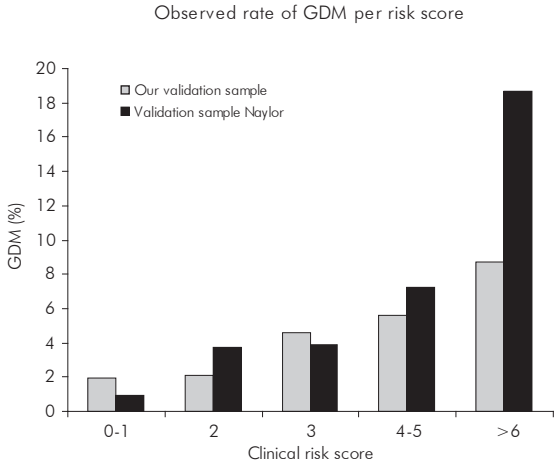


Figure 1. Correspondence between the risk score and the probability of GDM for Naylor's validation sample and our validation sample.

Table 3. Distribution of women across the predefined risk categories

	Risk category		
	Low	Intermediate	High
Our sample			
No. of women	311	593	362
No. of women with GDM (%)	6 (1.9)	20 (3.4)	21 (5.8)
Naylor's validation sample			
No. of women	544	606	421
No. of women with GDM (%)	5 (0.9)	23 (3.5)	41 (9.7)

prevalence of GDM increased with an increasing risk score in both samples. For most of the clinical risk scores, the prevalence of GDM in our sample was lower than the prevalence of GDM in Naylor's validation sample. The χ^2 goodness of fit test indicated a poor fit for the clinical scoring system in our sample ($\chi^2 = 8.89$, d.f. = 4, $p = 0.06$). After classifying women according to the predefined risk groups, the prevalence of GDM was 1.9%, 3.4% and 5.8% in the low, intermediate and high risk group respectively.

The performance of the screening strategy based on the clinical scoring system in our cohort is displayed in Table 4. If all women would be screened with the 50-g glucose challenge test (universal screening) the detection rate of GDM in our sample would be 68% (95% CI 55 - 80), with a corresponding false-positive rate of 10.8% (95% CI 10.4 - 11.4). If we would apply the screening strategy suggested by Naylor *et al.* to our sample, screening with the 50-g glucose challenge test could be omitted in 25% of the women. The detection rate in our cohort would consequently decrease to 64% (95% CI 50 - 76) with a corresponding false-positive rate of 12.6% (95% CI 12.1 - 13.1). Compared to universal screening, the decrease in detection rate was not statistically significant ($p = 0.48$). The false-positive rate of the screening strategy however, was significantly higher compared to universal screening ($p < 0.001$). In Naylor's derivation sample 35% of the women could refrain from screening by using the clinical scoring system. The detection

Table 4. Detection rate and false positive test results with universal screening and with the selective screening strategy in our sample. The screening test used was the 50-g glucose challenge test.

Strategy	No to be screened (%)	Detection Rate (n) [a]	P-value [b]	False-positive rate (n) [c]	P-value [b]
Usual Care [d]	1266	68.1% (32)		10.8% (132)	
Selective Screening [e]	955 (75.4%)	63.8% (30)	0.48	12.6% (153)	0.00

[a] values are based on a total of 47 true-positive and false-negative test results. [b] P-value for the comparison of selective screening with universal screening. [c] values are based on a total of 1219 false-positive and true-negative test results. [d] test 100% of the women with the 50-g glucose challenge test (threshold 7.8 mmol/L) [e] no screening if score 0-1, screening with the 50-g glucose challenge test (threshold 7.8 mmol/L) if score 2-3, screening with the 50-g glucose challenge test (threshold 7.1 mmol/L) if score >3

rate as well as the false-positive rate of the screening strategy in Naylor's derivation sample were significantly higher compared to our sample (detection rate 72.7%, p = 0.84; false positive rate 16.7%, p = 0.012).

Discussion

In this study we performed external validation of a clinical scoring system developed to perform selective screening for GDM. External validation is an essential step in the evaluation of a model or scoring system before it can be implemented in clinical practice,[6] as it assesses the performance of a developed model or scoring system in a different sample or in different circumstances than in which it was originally developed. In the present study we wanted to assess the validity of the scoring system when using the 75-g OGTT for the diagnosis of GDM instead of the 100-g OGTT. The validity of the scoring system in our sample could be considered as being unsatisfactory. The AUC of the scoring system in our sample was low, although comparable to the AUC of the scoring system in the original sample, and calibration indicated a poor fit for the individual risk score as well as for the predefined risk categories.

Naylor *et al.* developed their clinical scoring system by using converted odds ratios derived by multivariable logistic regression analysis. The odds ratios of all statistically significant variables were rounded to the nearest integer and added to develop the clinical risk score. No points were assigned for the reference categories of the variables. Criticism has been vented on this method of developing a clinical risk score. An odds ratio of 1.0 for all reference categories should be included in the scoring system, and when odds ratios are translated into a clinical scoring system, figures should be multiplied instead of added.[7] According to Naylor *et al.* they are aware of these flaws in their statistical framework; however they feel that the clinical scoring rule as they developed it does not lead to inferior performance of the scoring rule, whereas if official statistical procedures are followed, it would make the clinical scoring rule less understandable to practicing clinicians and unnecessarily cumbersome to use.[8]

A possible limitation of the present study is the limited number of women in the original sample in whom a reference test (OGTT) was performed. Since not all women underwent an OGTT to determine definitive GDM status, verification bias occurred. We corrected for this verification bias by means of multiple imputation. Imputing missing data is considered an eligible method to correct for verification bias and is preferred over complete case analysis, because in complete case analysis, deleting cases with missing values leads to a loss of statistical power and biased results.[9] Although it would have been preferable to perform an OGTT in all women, unfortunately this was not feasible in the original study,

due to the inconvenience this would have generated for the majority of the women. The poor calibration could be the result of poor fit of the scoring system to our sample, due to for example differences in sample characteristics or patient recruitment. Naylor *et al*. identified age as one of the risk factors for GDM. In our cohort there was no statistically significant association between age and GDM ($p = 0.14$). The association between older age and an increasing risk of GDM has been described in a number of studies.[10,11] A study by Coustan *et al*. showed an increasing incidence of GDM with increasing age.[10] In this study 56% of the women with GDM were younger than 30 years, which is consistent with the proportion in our sample (55.3%). In the present study we did not find an association between age and the risk of GDM when using the age categories defined by Naylor *et al*. When using age as a continuous variable, we still did not find a significant association.

Another reason for an unsatisfactory calibration could be differences in test protocol or analysis. Because we wanted to evaluate if the clinical scoring system was valid when using the 75-g OGTT as a diagnostic test, we used the 2 hour 75-g OGTT to diagnose GDM instead of the 3 hour 100-g OGTT. Since the original model was developed using the 3 hour 100-g OGTT as a diagnostic test, validation with the 2 hour 75-g alternative might have lead to inadequate estimation of the performance of the clinical scoring system in our sample. Application of a different reference test to validate the clinical scoring system might conceal the true origin of the poor calibration. In the present study we could not differentiate between poor fit due to application of a different reference test, or poor fit due to differences in sample characteristics. We have found however, that the scoring system has an unsatisfactory fit to our population when using the 75-g OGTT as a reference test. Santos-Ayarzagoitia described higher accuracy of the 100-g OGTT.[12] The results of the HAPO study however show that there is an association between the results of the 75-g OGTT and a number of important perinatal complications.[13] Since the 75-g OGTT nowadays is a frequently applied diagnostic test for GDM, in clinical practice as well as in a number of important studies, it is important that a clinical scoring system to perform selective screening is also applicable when using the 75-g OGTT.

Another explanation for the relatively poor performance of the scoring system is the slightly different gestational age at which the 50-g glucose challenge test was performed. In our sample the 50-g glucose challenge test was performed between 24 and 28 weeks of gestation, whereas in the study by Naylor *et al*. the 50-g glucose challenge test was performed at 25 to 27 weeks. In GDM insulin sensitivity decreases progressively with gestational age, leading to rising glucose values as pregnancy progresses. In line with this change in glucose tolerance, it is possible that women in our sample would have had different results on the 50-g glucose test, if the test was performed from 25 weeks to 27 weeks instead of 24 weeks to 28 weeks, leading to inaccurate estimates of the

performance of the scoring system. This would especially be the case if many women in our dataset would have a 50-g glucose challenge test result that is close to the predefined cutoff value of the 50-g glucose challenge test of 7.8 mmol/L. Since there were 73 women who had a 50 g glucose challenge test result ranging from 7.3 mmol/L to 7.8 mmol/L, of which 6 women had GDM. This false-negative rate might have been lower if women would have undergone the 50-g glucose challenge test later in pregnancy.

Although the performance of the scoring system was relatively poor in our sample, the performance of the screening strategy that was developed based on the scoring system was satisfactory. The detection rate of the selective screening strategy was comparable to the detection rate of universal screening in our cohort. By using the proposed screening strategies, screening with the 50-g glucose challenge test can be omitted in nearly 25% of the women in our sample, though with a higher false-positive rate and therefore at the cost of an increased rate of unnecessary performed diagnostic OGTTs of 1.8% (false-positive rate).

Some international expert groups recommend reduction of the upper limit of normal of the fasting venous plasma glucose of the OGTT from 7.0 mmol/L to 6.0 mmol/L, as this is considered to be more representative of the physiological changes in pregnancy.[14] Results from a large cohort study (HAPO study) show that there is a relation between the fasting value of the OGTT and the risk of a number of perinatal and maternal outcomes and complications.[13] Next to being more representative of the physiological changes in pregnancy, lowering the upper limit of normal fasting glucose values could also result in detecting more women at risk for complications. If these women are treated in a timely fashion, pregnancy outcome might be improved.

In conclusion, this study demonstrates that the discriminative capacity of the clinical scoring system developed by Naylor *et al*. in our sample was relatively poor and that the clinical scoring system estimated the risk of GDM only moderately using the 75-g OGTT as a diagnostic test for GDM. Performance of the screening strategy based on the scoring system however was still adequate, resulting in a reduction of rate of women needed to be screened of 25%, with a detection rate comparable to universal screening. Possibilities of another prediction model or a clinical scoring system for our population are worthwhile to explore since risk estimation of GDM was not optimal in our sample. A new prediction model or scoring system with additional or different prognostic factors or covariates could estimate the risk of GDM in our population more accurately, possibly improving the process of selective screening for GDM even further, leading to better patient care as well as to cost-effective management.

References

1. Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med 2005; 352:2477-2486.
2. American Diabetes Association. Gestational diabetes mellitus. Diabetes Care 2004; 27 Suppl 1: S88-S90.
3. Naylor CD, Sermer M, Chen E, Farine D. Selective screening for gestational diabetes mellitus. Toronto Trihospital Gestational Diabetes Project Investigators. N Engl J Med 1997; 337:1591-1596.
4. van Leeuwen M, Zweers EJ, Opmeer BC, van Ballegooie BE, ter Brugge HG, de Valk HW et al. Comparison of accuracy measures of two screening tests for gestational diabetes mellitus. Diabetes Care 2007; 2779-2784.
5. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and Classification of Diabetes Mellitus; Report of a WHO consultation. 1999.
6. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. N Engl J Med 1985; 313:793-799.
7. Harrell F. Regression coefficients and scoring rules. J Clin Epidemiol 1996; 49: 819.
8. Tu JV, Naylor CD. Clinical prediction rules. J Clin Epidemiol 1997; 50:743-744.
9. van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. J. Clin. Epidemiol 2006; 59:1102-1109.
10. Coustan DR, Nelson C, Carpenter MW, Carr SR, Rotondo L, Widness JA. Maternal age and screening for gestational diabetes: a population-based study. Obstet Gynecol 1989; 73:557-561.
11. Ogonowski J, Miazgowski T, Homa K, Celewicz Z, Kuczynska M. Low predictive value of traditional risk factors in identifying women at risk for gestational diabetes. Acta Obstet Gynecol Scand 2007: 86(10):1165-70.
12. Santos-Ayarzagoitia M, Salinas-Martinez AM, Villarreal-Perez JZ. Gestational diabetes: Validity of ADA and WHO diagnostic criteria using NDDG as the reference test. Diabetes Res Clin Pract 2006; 74:322-328.
13. Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR et al. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008; 358:1991-2002.
14. Brown CJ, Dawson A, Dodds R, Gamsu H, Gillmer M, Hall M et al. Report of the Pregnancy and Neonatal Care Group. Diabet Med 1996; 13:S43-S53.

Chapter 7

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

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

British Journal of Obstetrics and Gynaecology. 2010; 117(1):69-75

Abstract

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

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

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

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

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

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

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

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

Introduction

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

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

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

Methods

Patients and data

We used data from a prospective cohort study that compared the performance of the 50-g glucose challenge test and the random glucose test as screening tests for GDM.[18] All women with a singleton pregnancy who reported for prenatal care during a period of 2 years were invited to participate in this study. Women with known pre-gestational diabetes mellitus and women who were first seen after 20 weeks of gestation were excluded from the study. The following characteristics were obtained from all participating women at intake: obstetric history (parity, previous miscarriage, history of GDM, history of perinatal death), family history of diabetes mellitus (defined as a first- or second-degree relative with diabetes mellitus type I or II), ethnicity (self-reported), height and weight, age and smoking habits during pregnancy (categorised as smoking or non-smoking). The BMI before pregnancy was calculated as weight (kg)/[height (m)]^2.

Women underwent random glucose testing as well as a 50-g glucose challenge test. Both tests were performed once between the 24th and 28th week of gestation. If the random plasma glucose was higher than or equal to 6.8 mmol/l, or if the plasma glucose value at 1 hour after administration of 50 g glucose was higher than or equal to 7.8 mmol/l, a 2 hour, 75-g oral glucose tolerance test (OGTT) was performed within 1 week to confirm or rule out the presence of GDM (reference test). The OGTT was performed in the morning after a 12 hour overnight fast and after 3 days of minimal 150-200 g carbohydrate diet. Plasma glucose was determined before and 2 hours after the administration of a 75-g glucose-containing solution. GDM was considered to be present if the 2 hour venous plasma glucose value equalled or exceeded the cut-off value of 7.8 mmol/l, or if the fasting value was >7.0 mmol/l, according to the World Health Organisation (WHO) criteria.[19] If women with two negative screening tests were not tested with the reference test, some women with GDM (with a false-negative screening test result) would remain undetected, consequently generating verification bias. To estimate the proportion of diseased women who were not identified by the screening tests (false-negative fraction), in order to correct for verification bias, we performed an OGTT in a subset of women with two negative screening test results. Subsequently, we used multiple imputation to estimate the results of OGTT in all women in whom no OGTT was performed, based on the results of the two screening tests as well as on patient characteristics. Details of this procedure have been described elsewhere.[18] The original study was performed in two perinatal centres (Isala Clinics in Zwolle and the University Medical Centre in Utrecht). For the development of the prediction model, we used data from one centre (University Medical Centre in Utrecht).

Development of the prediction model

We used multiple logistic regression analysis to develop a statistical prediction model for GDM consisting of medical history and clinical risk indicators. For the development of the model, we evaluated the assumption of linearity in the logistic regression function for the continuous predictor variables age and BMI, using piecewise polynomials (splines) and visual inspection.[20] When the association was found to be non-linear, the variable was transformed to approach linearity. To determine the association between each predictor variable and the occurrence of GDM, we calculated univariable odds ratios (ORs), 95% confidence intervals (95% CIs) and P-values. Subsequently, we performed multivariable logistic regression analysis with a stepwise backwards selection procedure to construct the prediction model. Traditionally, a significance level of 5% in the univariable analysis is required for a variable to enter the multivariable logistic regression model. However, to avoid the erroneous exclusion of a potential relevant predictive variable, we increased the required significance level to enter the model to 30%.[21] A significance level of 20% was applied for a variable to stay in the model. Final model parameters were estimated using the SAS procedure MIANALYZE (multiple imputation procedure), which reflects uncertainty for imputed values using the slightly different estimates of the model parameters of the imputed datasets.

As the performance of prediction models is generally overestimated when applied in clinical practice (optimism), we adjusted the parameter estimates using a shrinkage factor λ, calculated as $\lambda = (\chi 2 - k)/\chi 2$, where $\chi 2$ is the likelihood ratio test and k is the number of covariates in the model.[22,23] All model parameters were uniformly shrunken with this shrinkage factor to adjust for optimism. The discriminative performance of the model was assessed by receiver operating characteristic (ROC) curve analysis and calculation of the area under the curve (AUC). An AUC of 0.5 indicates that the scoring system does no better than chance in discriminating between diseased and non-diseased women, whereas a scoring system with perfect discrimination would have an AUC of 1.0. Agreement between the predicted and observed probabilities was evaluated by plotting the mean predicted probabilities in ten risk groups (deciles) as calculated by the model, against the observed proportion of women with GDM in these groups (calibration). The goodness of fit of the model was evaluated with the Hosmer and Lemeshow test statistic.[24]

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

Results

Patients and data

We used data from 995 women who were included in the original cohort study.[18] Random glucose testing was performed in 995 women (100%). The 50-g glucose challenge test was performed in 978 women (98.3%). Thirty-one of the 995 women (3.2%) had an abnormal random glucose value, and 99 of the 978 women (10.1%) had an abnormal result on the 50-g glucose challenge test; 114 women (11.5%) had at least one abnormal test result (random glucose test or 50-g glucose challenge test, or both). In 16 women (1.6%), both test results were suspect for GDM. Of the 114 women with at least one abnormal screening test result, 37 women did not consent to undergo an OGTT. Oral glucose tolerance test was performed in 122 women (12.3%), either because of at least one abnormal screening test result (n = 93) or because the women were part of the subgroup in which OGTT was performed irrespective of two negative screening test results (n=29). Of these 122 women, 22 were diagnosed with GDM and 100 had no GDM. In 873 women, initially no OGTT was performed. The procedure to correct for verification bias indicated that two of these women had GDM, resulting in a total of 24 of 995 women (2.4%) diagnosed with GDM after correction for verification bias.

Development of the clinical prediction model

The assumption of linearity was satisfied in age, but not in BMI. The risk of GDM increased with an increasing BMI between 22 and 30 kg/m^2, but below 22 kg/m^2 and above 30 kg/m^2 the risk of GDM did not alter any further (Figure 1). We therefore performed a

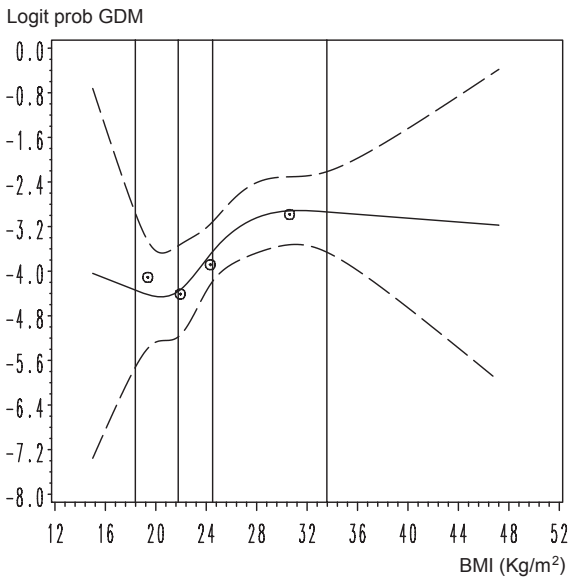


Figure 1. Association between BMI (kg/m2) and the (logit) probability of gestational diabetes mellitus (GDM) with 95% CI. BMI, body mass index.

simple transformation for BMI. A BMI below 22 kg/m^2 was rounded up to 22 kg/m^2, and a BMI above 30 kg/m^2 was rounded down to 30 kg/m^2, for which the association between BMI and the probability of GDM was linear. We planned to evaluate the associations between various ethnicity categories (n = 8) and the occurrence of GDM. However, as the number of women in the various ethnicity categories was too small to demonstrate statistically different associations, in our analyses we only differentiated between Caucasian and non-Caucasian ethnicity.

The results of the univariable and multivariable analyses are summarised in Table 1. The shrinkage coefficient was 0.73, indicating that the performance of the model was overestimated by 27%. The multivariable analysis (after shrinkage) showed that non-Caucasian ethnicity (OR 2.3 (95% CI 1.2 - 4.6)), a family history of diabetes (OR 1.8 (95% CI 0.9 - 3.3)), history of GDM (OR 0.5 (95% CI 0.3 - 1.0) for women without a history of GDM compared to nullipara, and OR 1.6 (95% CI 0.3 - 8.3) for women with a history of GDM compared to nullipara) and BMI (per kg/m^2) (OR 1.14 (95% CI 1.04 - 1.25)) increased the risk of GDM. The probability of GDM in our population can be calculated using the formula representing the logistic regression model (after shrinkage): probability of GDM = $1/[1 + \exp(-\beta)]$, in which β is calculated as $[-6.1 + (0.83 \times$ non-Caucasian ethnicity) $+ (0.57 \times$ positive family history of diabetes mellitus) $- (0.67 \times$ multipara without history of GDM) $+ (0.5 \times$ multipara with history of GDM) $+ (0.13$

Table 1. Baseline characteristics of our cohort and results of uni- and multivariable analysis.

	Univariable Analysis			Multivariable Analysis		
	OR	95% CI	P	OR	95% CI	P
Age (per year)	1.0	0.93 - 1.02	0.87			
BMI (per kg/m^2)	1.2	1.06 - 1.37	0.0036	1.14	1.03 - 1.26	0.009
Ethnicity non-Caucasian	3.7	1.5 - 9.1	0.0036	2.3	1.2 - 4.6	0.02
Family history of diabetes	2.5	1.1 - 5.8	0.03	1.8	0.9 - 3.3	0.08
Smoking	1.2	0.4 - 3.2	0.74			
Previous miscarriage	1.1	0.7 - 1.7	0.59			
History of GDM						
Nullipara	1.0			1.0		
Multipara without history of GDM	0.5	0.2 - 1.3	0.15	0.5	0.3 - 1.0	0.05
Multipara with history of GDM	3.6	0.4 - 30.1	0.23	1.6	0.3 - 8.3	0.55
History of perinatal death						
Nullipara	1.0					
Multipara without perinatal death	0.6	0.2 - 1.5	0.27			
Multipara with perinatal death	0.7	0.1 - 5.4	0.71			

BMI, body mass index; CI, confidence interval; GDM, gestational diabetes mellitus; OR, odds ratio. 95% CI = 95% confidence interval; P = p value. Values represent the number of women (%) unless otherwise indicated. ORs in the multivariable analysis are corrected for over fit with the shrinkage procedure

Table 2. Clinical consequences of various strategies.

Screening strategy	Women tested with OGTT (n (%))	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	OGTT to diagnose 1 case of GDM (n)
Universal testing with OGTT	995 (100)	100	100	42 (995 / 24)
OGTT if probability of GDM ≥ 2.0%	428 (43.0)	75.0 (55.4 - 88.0)	57.8 (57.3 - 58.1)	24 (428 / 18)
OGTT if probability of GDM ≥ 4.0%	124 (12.5)	45.8 (28.2 - 64.5)	88.4 (87.9 - 88.8)	11 (124 / 11)

GDM = gestational diabetes mellitus. OGTT = 75-g oral glucose tolerance test. The number of women in the sample was 995. The number of women with GDM was 24.

× BMI)]. The mean AUC of the ROC curves from the ten multiple imputed datasets was 0.77 (95% CI 0.69-0.85), demonstrating a reasonable capacity to discriminate between women with and without GDM. The P-value for the Hosmer and Lemeshow test statistic was 0.25, indicating adequate agreement between the mean predicted probabilities and the observed probability of GDM (calibration).

Clinical consequences

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

Discussion

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

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

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

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

Table 3. Comparison of the scoring systems for the risk of gestational diabetes mellitus (GDM)

Screening strategy	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%) (95% CI)
Threshold Caliskan [15]			
≥ 1 point	100 (78.9 - 100)	32.1 (31.4 - 32.1)	4.8 (3.8 - 4.8)
≥ 2 points	85.7 (60.6 - 96.0)	64.5 (63.6 - 64.8)	7.6 (5.4 - 8.6)
≥ 3 points	57.1 (33.2 - 78.2)	86.3 (85.5 - 87.0)	12.5 (7.3 - 17.1)
≥ 4 points	35.7 (17.4 - 56.4)	97.8 (97.2 - 98.5)	35.7 (17.4 - 56.4)
≥ 5 points	7.1 (1.4 - 12.9)	99.6 (99.8 - 1.00)	50.0 (9.5 - 90.4)
Threshold Naylor [16]			
> 1 point	92.8 (84.3 - 96.9)	35.9 (35.5 - 36.1)	6.2 (5.7 - 6.5)
≥ 2 points	75.4 (64.3 - 83.9)	56.5 (56.0 - 56.9)	7.4 (6.3 - 8.2)
≥ 3 points	59.4 (48.0 - 70.0)	74.4 (74.2 - 75.2)	9.7 (7.9 - 11.5)
≥ 5 points	24.6 (16.4 - 34.7)	95.1 (94.7 - 95.5)	18.7 (12.4 - 26.3)
Threshold present study			
Prediction of 2%	75.0 (55.4 - 88.0)	57.8 (57.3 - 58.1)	4.2 (3.1 - 4.9)
Prediction of 4%	45.8 (28.2 - 64.5)	88.4 (87.9 - 88.8)	8.9 (5.5 - 12.5)

CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio; OGTT, 75-g oral glucose tolerance test. The accuracy measures of the scoring systems of Caliskan *et al*.[15] and Naylor *et al*.[16] were calculated from the original articles. Presentation of the results in the studies by Caliskan *et al*.[15] and Naylor *et al*.[16] differ from presentation of the results in the present study. The percentages and points are not directly comparable.

over complete case analysis.[25] By using multiple imputation instead of single imputation, the uncertainty regarding imputed values is statistically incorporated, resulting in more accurate confidence intervals.[25]

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

To use the prediction model in practice, ideally it should be electronically available to the clinician, preferably in an electronic patient file, where the probability of GDM is

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

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

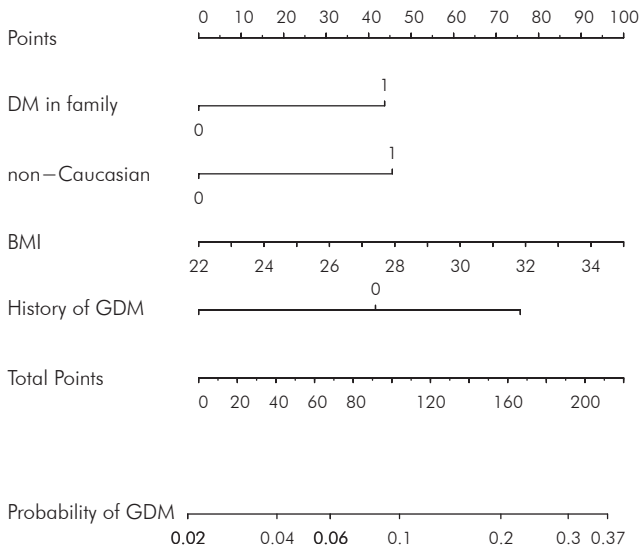


Figure 2. Nomogram to estimate the probability of gestational diabetes mellitus (GDM). BMI, body mass index.

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

Conclusion

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

References

1. King H. Epidemiology of glucose intolerance and gestational diabetes in women of childbearing age. Diabetes Care 1998; 21(Suppl. 2):B9-13.
2. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. Diabetes Care 2002; 25:1862-8.
3. Heineman MJ, Bleker OP, Evers JHL, Heinz APM. Obstetrics and Gynaecology. Gezondheidszorg: Elsevier, 2004.
4. Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med 2005; 352:2477-86.
5. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2009; 32(Suppl. 1):S62-7.
6. Alberico S, Strazzanti C, De SD, De SF, Lenardon P, Bernardon M, et al. Gestational diabetes: universal or selective screening? J Matern Fetal Neonatal Med 2004; 16:331-7.
7. Cosson E, Benchimol M, Carbillon L, Pharisien I, Paries J, Valensi P, et al. Universal rather than selective screening for gestational diabetes mellitus may improve fetal outcomes. Diabetes Metab 2006; 32:140-6.
8. Griffin ME, Coffey M, Johnson H, Scanlon P, Foley M, Stronge J, et al. Universal vs. risk factor-based screening for gestational diabetes mellitus: detection rates, gestation at diagnosis and outcome. Diabet Med 2000; 17:26-32.
9. Di Cianni G, Volpe L, Lencioni C, Miccoli R, Cuccuru I, Ghio A, et al. Prevalence and risk factors for gestational diabetes assessed by universal screening. Diabetes Res Clin Pract 2003; 62:131-7.
10. Jensen DM, Molsted-Pedersen L, Beck-Nielsen H, Westergaard JG, Ovesen P, Damm P. Screening for gestational diabetes mellitus by a model based on risk indicators: a prospective study. Am J Obstet Gynecol 2003; 189:1383-8.
11. Keshavarz M, Cheung NW, Babaee GR, Moghadam HK, Ajami ME, Shariati M. Gestational diabetes in Iran: incidence, risk factors and pregnancy outcomes. Diabetes Res Clin Pract 2005; 69:279-86.
12. Ogonowski J, Miazgowski T, Homa K, Celewicz Z, Kuczynska M. Low predictive value of traditional risk factors in identifying women at risk for gestational diabetes. Acta Obstet Gynecol Scand 2007; 86:1165-70.
13. Solomon CG, Willett WC, Carey VJ, Rich-Edwards J, Hunter DJ, Colditz GA, et al. A prospective study of pregravid determinants of gestational diabetes mellitus. J Am Med Assoc 1997; 278:1078-83.
14. Sunsaneevithayakul P, Boriboohirunsarn D, Sutanthavibul A, Ruangvutilert P, Kanokpongsakdi S, Singkiratana D, et al. Risk factor-based selective screening program for gestational diabetes mellitus in Siriraj Hospital: result from clinical practice guideline. J Med Assoc Thai 2003; 86:708-14.
15. Caliskan E, Kayikcioglu F, Ozturk N, Koc S, Haberal A. A population-based risk factor scoring will decrease unnecessary testing for the diagnosis of gestational diabetes mellitus. Acta Obstet Gynecol Scand 2004; 83:524-30.
16. Naylor CD, Sermer M, Chen E, Farine D. Selective screening for gestational diabetes mellitus. Toronto Trihospital Gestational Diabetes Project Investigators. N Engl J Med 1997; 337:1591-6.
17. Sacks DA, Abu-Fadil S, Karten GJ, Forsythe AB, Hackett JR. Screening for gestational diabetes with the one-hour 50-g glucose test. Obstet Gynecol 1987; 70:89-93.
18. van Leeuwen M, Zweers EJ, Opmeer BC, van Ballegooie E, ter Brugge HG, de Valk HW, et al. Comparison of accuracy measures of two screening tests for gestational diabetes mellitus. Diabetes Care 2007; 30:2779-84.

19. Report of a World Health Organization (WHO) Consultation. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications. Part 1. Geneva: WHO, 1999.
20. Harrell FE Jr, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. J Natl Cancer Inst 1988; 80:1198-202.
21. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med 2000; 19:1059-79.
22. Steyerberg EW. Clinical Prediction Models. New York: Springer Verlag, 2009.
23. Harrell FE Jr. Regression Modeling Strategies. New York: Springer Verlag, 2001.
24. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med 1997; 16:965-80.
25. van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. J Clin Epidemiol 2006; 59:1102-9.

Chapter 8

Gestational diabetes mellitus: treatment reduces the risk of complications

Marsha van Leeuwen
Suzanne M. Prins
Harold W. de Valk
Inge M. Evers
Gerard H.A. Visser
Ben Willem J. Mol

Adapted from Nederlands Tijdschrift voor Geneeskunde. 2011; 155:A2291

Abstract

Recent studies show that higher blood glucose values after a 75-g oral glucose tolerance test in pregnancy are associated with higher rates of perinatal and maternal complications. Treatment of gestational diabetes mellitus (GDM) (or hyperglycemia in pregnancy) reduces the risk of complications. GDM is an asymptomatic condition. Screening is the only strategy to diagnose GDM in time, in order to provide treatment. Until recently, there was no uniformity concerning diagnostic strategy and treatment of GDM in the Netherlands, possibly due to lack of evidence on the risk of complications and the effectiveness of treatment. Results of several recent studies show that early detection and treatment of GDM are effective. By means of a more active screening and treatment policy it should be possible to reduce the perinatal and maternal complications as a result of GDM.

Introduction

Gestational diabetes mellitus (GDM) is defined as carbohydrate intolerance of variable severity with onset or first recognition during pregnancy and includes carbohydrate intolerance first developing in pregnancy, as well as pre-existing diabetes mellitus that has not been recognized before.[1,2] Throughout the last years results of three large trials have been published that provide more clarity on the risk of complications and the effect of treatment of treatment of GDM. In this article we discuss the results of these trials. Furthermore we present the results of a national survey that we performed to determine current practices on screening and diagnostics of GDM in the Netherlands. We will discuss the results of survey in view of the recommendations from the renewed guideline "Diabetes and Pregnancy" issued by the Dutch Society of Obstetrics and Gynaecology (NVOG).

Criteria for GDM

Worldwide, and over the years there have been many different criteria for GDM all reflecting carbohydrate intolerance, albeit at different levels. The many different criteria preclude a sound comparison of research findings and extrapolation of study results. The lack of uniform criteria for GDM is amongst others founded by the fact that the original criteria for GDM established by O'Sullivan and Mahan in 1964, were initially selected to identify women at risk for developing diabetes mellitus (type 2) in the future and did not reflect the risk for complications during pregnancy and delivery.[3,4] In recent years, focus has been directed more on perinatal and short-term maternal outcomes.

Pathophysiology of GDM

Placental hormones produced in pregnancy hamper normal carbohydrate metabolism. Corticotropin-releasing hormone, progesterone and human placental lactogen interfere with insulin receptors situated on various cells of the human body, making these cells less sensitive for insulin, resulting in relative insulin deficiency. To maintain maternal blood glucose levels within normal range, production of insulin by the beta-cells of the pancreas is increased. If this compensating mechanism is insufficient, GDM may occur.

Prevalence of GDM

The prevalence of GDM has increased over the last years and is estimated to be 2 to 9% depending on the population studied and the criteria for GDM that are applied.[5,6] The prevalence of GDM in the Netherlands is estimated to be 2 to 4%. The prevalence of GDM is rising, mainly due to the rising epidemic of overweight and obesity and changes in lifestyle in developed countries.[5]

Complications of GDM and treatment

Pre-existing diabetes mellitus type 1 and 2 are associated with maternal complications and adverse perinatal outcome.[7-9] Various studies described that GDM is associated

with pregnancy complications too.[10,11] In addition, women with GDM are at increased risk for diabetes mellitus type 2 in later life. Until recently however, exact risks associated with hyperglycemia in pregnancy less severe than overt diabetes mellitus were undecided. Moreover, the degree of carbohydrate intolerance at which the risk of specific complications increases was unspecified. In addition, before 2005 there was no evidence that treatment of GDM would reduce the risks of pregnancy complications.[12,13] This has led to international but also national variety in the clinical practice of screening, diagnostics and treatment of GDM. We will discuss the results of three trials that have provided more clarity on the risk of complications and the effect of treatment of treatment of GDM.

Risk of complications

The 'Hyperglycemia and adverse pregnancy outcome' (HAPO)-study is an international cohort study that investigated the association between maternal blood glucose levels in pregnancy and the risk of adverse perinatal and maternal outcome.[14] Aim of the study was to evaluate the association between the fasting, 1 hour and 2 hour plasma glucose value of the 75-g oral glucose tolerance test (OGTT) and various perinatal and maternal outcomes of pregnancy. Over 25 505 women with a singleton pregnancy underwent a 75-g oral glucose tolerance test (OGTT) between 24 and 32 weeks of gestation. Women with pre-existing diabetes mellitus were excluded. Primary outcomes of the study were birth weight above the 90th percentile for gestational age, primary caesarean delivery, clinical neonatal hypoglycemia, and cord-blood serum C-peptide level above the 90th percentile (fetal hyperinsulinemia). Secondary outcomes were premature delivery (< 37 weeks), shoulder dystocia or birth injury, need for intensive neonatal care, hyperbilirubinemia, and pre-eclampsia. Outcomes were adjusted for pre-specified confounders. Results from the HAPO study show positive associations between increasing levels of fasting, 1-hour, and 2-hour plasma glucose of the 75-g OGTT and the risk of virtually all perinatal and maternal primary and secondary outcomes after adjustment for confounders. Higher maternal blood glucose levels are significantly associated with adverse outcomes, even with glucose levels that are within the range that is considered as non-diabetic.

Effect of treatment of GDM

Two large randomised controlled trials evaluated the effect of treatment of GDM. The aim of the 'Australian carbohydrate intolerance study in pregnant women' (ACHOIS) was to evaluate if treatment of GDM reduced the risk of perinatal complications.[15] Pregnant women with risk factors for GDM or women with an abnormal result of the 50-g glucose challenge test underwent a 75-g OGTT between 24 and 34 weeks of gestation. GDM was diagnosed if the venous plasma glucose level was < 7.8 mmol/L after overnight fasting and 7.8-11.0 mmol/l two hours after the OGTT. Women diagnosed with GDM were randomised (intervention group and control group). Women who were assigned to

the intervention group were frequently seen by a physician, performed self-monitoring of blood glucose values four times a day, received individualised dietary advice and insulin therapy if necessary. Women in the routine care group received regular obstetric care (local protocol). Women in the routine care group and their physicians were unaware of the diagnosis of glucose intolerance of pregnancy. A proportion of the women with normal OGTT results were assigned to the routine care group to maintain blinding. Thousand women were included. The rate of serious perinatal outcomes (composite outcome: death, shoulder dystocia, bone fracture, and nerve palsy) among infants was significantly lower in the intervention group than the routine care group (1 vs. 4%, P=0.01). The number needed to treat to prevent one serious outcome was 34 (95% confidence interval 20 - 103). Induction of labour was significantly more common in the intervention group and more infants born to women in the intervention group were admitted to the neonatal nursery. There was no difference in the rate of caeserean sections (emergency or planned). Mean birth weight of infants born to women in the intervention group was lower, and they were born at an earlier gestational age. Fewer women in the intervention group received diagnosis of pre-eclampsia. Scores on mental health status were in favor of women in the intervention group. The ACHOIS shows that treatment of GDM (by means of dietary advice, blood glucose monitoring and insulin therapy if required) reduces the rate of perinatal complications. This also leads to more induction of labour in women with GDM and admittance of more infants to the neonatal nursery, but not to higher rates of caeserean sections.

The second study on this subject aimed to evaluate if treatment of mild GDM would lead to fewer perinatal and obstetric complications.[16] Women with a singleton pregnancy and an abnormal result on the 50-g glucose challenge test (glucose value of 7.5-11.1 mmol/l) underwent a 100-g OGTT between 24 and 31 weeks of gestation. "Mild GDM" was diagnosed based on the results of the 100-g OGTT. Women with mild GDM were randomised between an intervention group and a control group. Treatment in the intervention group comprised regular self-monitoring of glucose values, dietary advice and treatment with insulin if required. Women in the control group received standard obstetric care (local protocol). A proportion of the women with normal OGTT results were assigned to the routine care group to maintain blinding. Primary outcome was a composite outcome of stillbirth or perinatal death and neonatal complications, including neonatal hyperbilirubinemia, hypoglycemia, hyperinsulinemia, and birth trauma. There were 958 women included in the study. There was no significant difference between groups in incidence of the composite outcome (stillbirth or perinatal death and neonatal complications, including neonatal hyperbilirubinemia, hypoglycemia, hyperinsulinemia, and birth trauma). However, treatment of mild GDM was associated with reduced rates of several pre-specified secondary outcomes, e.g. shoulder dystocia, caesarean delivery, frequency of large-for-gestational-age infants and birth weight > 4000 g. Furthermore

treatment of GDM was associated with reduced rates of pre-eclampsia and gestational hypertension. Although women included in this trial had a milder degree of glucose intolerance than women included in the ACHOIS[13], results of this study show that treatment of mild GDM does not lead to a reduction of perinatal mortality, neonatal hypoglycemia, hyperbilirubinemia, hyperinsulinemia, and birth trauma. However, treatment of mild GDM does lead to reduced rates of shoulder dystocia, caesarean delivery, frequency of large-for-gestational- age infants, birth weight > 4000 g and gestational hypertension and pre-eclampsia.

Summary of literature

The studies described above show that higher glucose values in pregnancy are associated with a number of important perinatal and maternal outcomes. Treatment of (mild) GDM reduces the rate of serious perinatal complications (composite outcome: death, shoulder dystocia, bone fracture, and nerve palsy), pre-eclampsia and delivery of large-for-gestational- age infants and birth weight > 4000 g.

Screening policy in the Netherlands

In 2006 the Dutch multidisciplinary guideline "Diabetes and Pregnancy" was published. The guideline recommended to perform either a random glucose measurement or a fasting glucose measurement at the first pregnancy check up in order to detect hyperglycemia that already exists early in pregnancy (unknown pre-existing diabetes mellitus). Screening in the second trimester was not recommended because at the time there was insufficient evidence for screening in the second trimester to be effective.[17] In view of the results of the above-mentioned trials the latter recommendation was reconsidered. In the guideline of the Dutch Society of Obstetrics and Gynaecologists (NVOG) issued in 2010 screening in the second trimester is recommended.[18] It is recommended to perform screening in the first and second trimester in women with risk factors for GDM (history of GDM, BMI $>30 kg/m^2$, previous macrosomic baby, first degree relative with DM, certain ethnicities, history of (unexplained) perinatal death, polycystic ovary syndrome). The screening test recommended for the first trimester is a random or fasting glucose measurement. The test recommended in the second trimester is the 75-g OGTT. [18] If women had gestational diabetes in a previous pregnancy it is recommended to perform a 75-g OGTT in the 16^{th} week of pregnancy.

Survey

Because we suspected a large variability in the screening, diagnosis, and management of women with GDM, we performed a national survey to determine current policy on GDM in the Netherlands. We performed this survey before 2010, so in the period before the new Dutch NVOG guideline was issued. The survey was web-based (www.questionpro.com) and comprised questions on screening and treatment of GDM. For all Dutch hospitals with an obstetric department (n=93) one gynaecologist received an invitation by email to participate in the survey. Furthermore we randomly invited 129 midwives

Table 1. Policy on screening in the Netherlands (midwives and gynaecologists) before 2010.

	Gynaecologists n (%)	Midwives n (%)	Total n (%)
Screening 1st trimester			
Yes	46 (67.6)	75 (81.5)	121 (75.6)
No	22 (32.4)	17 (18.5)	39 (24.4)
Screening 2nd trimester			
Yes	36 (52.9)	60 (65.2)	96 (60.0)
No	32 (47.1)	32 (34.8)	64 (40.0)
Screening 2nd trimester			
All women	15 (41.7)	21 (35.0)	36 (37.5)
Women with risk factors	21 (58.3)	39 (65.0)	60 (62.5)
Methods of screening in 2nd trimester			
Fasting glucose measurement	17 (47.2)	28 (46.7)	45 (46.9)
Random glucose measurement	29 (80.6)	45 (75.0)	74 (77.1)
Challenge test with 50 g glucose	9 (25.0)	20 (33.3)	29 (30.2)
Universal OGTT	8 (22.2)	0 (0)	8 (8.3)
Breakfast or lunch test	15 (41.7)	19 (31.7)	34 (35.4)
Day curve	17 (47.2)	26 (43.3)	43 (44.8)
Risk factors	32 (88.9)	51 (85.0)	83 (86.5)
Most frequently used method in 2nd trimester			
Fasting glucose measurement	1 (2.8)	3 (5.0)	4 (4.2)
Random glucose measurement	18 (50.0)	26 (43.3)	44 (45.8)
Challenge test with 50 g glucose	2 (5.6)	1 (1.7)	3 (3.1)
Universal OGTT	0 (0)	0 (0)	0 (0)
Breakfast or lunch test	6 (16.7)	8 (13.3)	14 (14.6)
Day curve	3 (8.3)	9 (15.0)	12 (12.5)
Risk factors	6 (16.7)	13 (21.7)	19 (19.8)

from different practices to participate in the survey. Response rates were 73% (68/93) and 71% (92/129) for gynaecologists and midwives respectively. Results of the survey are shown in Tables 1 and 2. The majority of the gynaecologists and midwives performed screening in the first trimester (68% (46/68) and 82% (75/92) respectively). Screening in the second trimester was performed by 53% (36/68) of the gynaecologists and 65% (36/68) of the midwives, mostly screening was performed in women with risk factors for GDM. There was a large variety in screening strategies. Random glucose measurement was the most frequently used test for screening in the second trimester. Most frequently used test to diagnose GDM was a "lunch" test (measuring glucose values one and two hours after lunch) (43% (29/68)), followed by the 75-g OGTT (31% (21/68)) and a series of multiple random measurements on one day (19% (13/68)).

Table 2. Diagnostic testing before 2010.

	Gynaecologists n (%)
Diagnostic test	
If screening test is abnormal	43 (63.2)
If symptoms present	37 (54.4)
If risk factors present	34 (50.0)
Universal	8 (11.7)
Most frequently used diagnostic test	
OGTT 75 gram	21 (30.9)
OGTT 100 gram	4 (5.9)
Breakfast / lunch test	29 (42.6)
Day curve	13 (19.1)

Discussion

GDM is associated with higher levels of maternal and neonatal morbidity and treatment of GDM reduces the rate of complications.[14-16] Timely detection of GDM is important, but is also difficult since GDM often is asymptomatic. A way to identify women with GDM is screening. The results of a survey amongst Dutch gynaecologists and midwives show that before introduction of the guideline "Diabetes and Pregnancy" issued by the Dutch Society of Obstetrics and Gynaecology (NVOG) in 2010 there was large variety in the policy on screening and diagnosis of GDM. Most frequently applied screening test in the second trimester of pregnancy was random glucose testing. The random glucose test has moderate reproducibility and accuracy[19] and therefore should not be used as a screening test in the second trimester of pregnancy. Nearly 2/3 of gynaecologists performed a lunch test or a series of multiple measurements on one day to diagnose GDM. Reproducibility and accuracy of these tests for GDM however are unknown. The HAPO study as well as the two intervention studies that were discussed in this article used the OGTT to diagnose GDM.[14-16] Although critics consider the OGTT to be not physiologic and therefore unreliable, reproducibility is acceptable (75-79%).[20,21]

The use of inaccurate and inconsistent test strategies may lead to under-diagnosis and suboptimal treatment of GDM leading to potentially avoidable complications. The current Dutch guideline on diabetes and pregnancy recommends screening of at least all women with risk factors for GDM in the 1st trimester of pregnancy, by means of fasting or random glucose measurement. For women with risk factors 2nd trimester screening with a 75-g OGTT is recommended. For women with GDM in a previous pregnancy a 75-g OGTT at 16 weeks of gestation is recommended, followed by an OGTT between 24 and 28 weeks of pregnancy if the first OGTT is normal.[18] It is unknown which part of the midwives and gynaecologists follow these recommendations from the renewed guideline.

Conclusion

GDM is associated with higher levels of maternal and neonatal morbidity and treatment of GDM reduces the rate of complications.[14-16] Further studies on costs and effects of various screening strategies for GDM need to clarify the optimal strategy to identify women with GDM in order to reduce the number of perinatal and maternal adverse outcomes. Until more evidence on costs and effects of various screening strategies is generated, the pathway and treatment strategy as recommended by the guideline "Diabetes and Pregnancy" of the Dutch Society of Obstetrics and Gynaecologists (NVOG) based on the best available evidence should be followed

Literature

1. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and Classification of Diabetes Mellitus; Report of a WHO consultation. Diabet Med 1998; 15:539-53.
2. American Diabetes Association. Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 2009; 32(Supp 1):S62-S67.
3. O'Sullivan JB, Mahan CM. Criteria for oral glucose tolerance test in pregnancy. Diabetes 1964; 13:278-85.
4. Metzger BE, Buchanan TA, Coustan DR, de Leiva A, Dunger DB, Hadden DR et al. Summary and recommendations of the Fifth International Workshop-Conference on Gestational Diabetes Mellitus. Diabetes Care 2007; 30(Suppl. 2):S251-S260.
5. Getahun D, Nath C, Ananth CV, Chavez MR, Smulian JC. Gestational diabetes in the United States: temporal trends 1989 through 2004. Am J Obstet Gynecol 2008; 198:525.e1-5.
6. King H. Epidemiology of glucose intolerance and gestational diabetes in women of childbearing age. Diabetes Care 1998; 21(Suppl 2):B9-B13.
7. Cundy T, Gamble G, Townend K, Henley PG, MacPherson P, Roberts AB. Perinatal mortality in Type 2 diabetes mellitus. Diabet Med 2000; 17:33-9.
8. Evers IM, de Valk HW, Visser GH. Risk of complications of pregnancy in women with type 1 diabetes: nationwide prospective study in the Netherlands. BMJ 2004; 328:915.
9. Cheung NW, McElduff A, Ross GP. Type 2 diabetes in pregnancy: a wolf in sheep's clothing. Aust N Z J Obstet Gynaecol 2005; 45:479-83.
10. Sermer M, Naylor CD, Gare DJ, Kenshole AB, Ritchie JW, Farine D et al. Impact of increasing carbohydrate intolerance on maternal-fetal outcomes in 3637 women without gestational diabetes. The Toronto Tri-Hospital Gestational Diabetes Project. Am J Obstet Gynecol 1995; 173:146-56.
11. Yang X, Hsu-Hage B, Zhang H, Zhang C, Zhang Y, Zhang C. Women with impaired glucose tolerance during pregnancy have significantly poor pregnancy outcomes. Diabetes Care 2002; 25:1619-24.
12. Jarrett RJ. Gestational diabetes mellitus. BMJ 1993; 306:1065.
13. Greenberg JA. Gestational diabetes: what's the problem? Am J Obstet Gynecol 2006; 194:299.
14. Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR et al; HAPO Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008; 358:1991-2002.
15. Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS; Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) Trial Group. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med 2005; 352:2477-86.
16. Landon MB, Spong CY, Thom E, Carpenter MW, Ramin SM, Casey B et al. A multicenter, randomized trial of treatment for mild gestational diabetes. N Engl J Med 2009; 361:1339-48.
17. Nederlandse Internisten Vereniging. Richtlijn 'Diabetes en zwangerschap'. 2006. https://caeusr-9jmbgskxm6l4ke.sec.amc.nl/home/richtlijnen/niv/niv/diabetes-en-zwangerschap.
18. Nederlandse Vereniging voor Obstetrie en Gynaecologie. Richtlijn 'Diabetes en zwangerschap' Versie 2.0; 2010. https://caelrjg067l2zojgkch7gj.sec.amc.nl/richtlijn/doc/index.php?type=save&richtlijn_id=863.
19. van Leeuwen M, Zweers EJ, Opmeer BC, van Ballegooie E, ter Brugge HG, de Valk HW et al. Comparison of accuracy measures of two screening tests for gestational diabetes mellitus. Diabetes Care 2007; 30:2779-84.
20. Catalano PM, Avallone DA, Drago NM, Amini SB. Reproducibility of the oral glucose tolerance test in pregnant women. Am J Obstet Gynecol 1993; 169:874-81.
21. Harlass FE, Brady K, Read JA. Reproducibility of the oral glucose tolerance test in pregnancy. Am J Obstet Gynecol 1991; 164:564-8.

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,[13] 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.[14] 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 3 537 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.[15] 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.[16] 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 witch 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.[17] 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.[18,19] 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.[20] 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.[20] 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).[2] 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.[20] 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 ceasarean section, clinical neonatal hypoglycemia, and cord-blood serum C-peptide level > 90th percentile).[20]

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 with 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,[1,2,10-12] 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.[18] 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.[19] 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.[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.[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.[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.[23] Since treatment strategies for obesity have rarely been successful,[24] 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.[2] 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.[2] 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.[1,10] 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 [25,26] to explore if this is an accurate method to select women who will benefit form 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 OGTT for all pregnant women reduced GDM related complications considerable for an acceptable cost-effectiveness ratio. So, if universal screening with an OGTT 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.[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 OGTT in general is considered to be an unpleasant test for pregnant women,[25,26] we do not have information on women's preferences regarding the OGTT 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 OGTT 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 OGTT 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 OGTT 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.[15] 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

1. Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS; Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) Trial Group. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med 2005; 352(24):2477-86.
2. Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR et al. HAPO Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008; 358(19):1991-2002.
3. Wendland EM, Torloni MR, Falavigna M, Trujillo J, Dode MA, Campos MA et al. Gestational diabetes and pregnancy outcomes-a systematic review of the World Health Organization (WHO) and the International Association of Diabetes in Pregnancy Study Groups (IADPSG) diagnostic criteria. BMC Pregnancy and Childbirth 2012; 12:23
4. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. Diabetes Care 2002; 25(10):1862-8.
5. Yogev Y, Visser GH. Obesity, gestational diabetes and pregnancy outcome. Semin Fetal Neonatal Med 2009; 14:77-84.
6. Vohr BR, Boney CM. Gestational diabetes: the forerunner for the development of maternal and childhood obesity and metabolic syndrome? J Matern Fetal Neonatal Med 2008; 21:149-57.
7. Hillier TA, Pedula KL, Schmidt MM, Mullen JA, Charles MA, Pettitt DJ. Childhood obesity and metabolic imprinting: the ongoing effects of maternal hyperglycemia. Diabetes Care 2007; 30(9):2287-92.
8. Hunt KJ, Schuller KL. The increasing prevalence of diabetes in pregnancy. Obstet Gynecol Clin North Am 2007; 34(2):173-99.
9. Wilson JM, Jungner YG. Principles and practice of mass screening for disease. Bol Oficina Sanit Panam 1968; 65(4):281-393.
10. Landon MB, Spong CY, Thom E, Carpenter MW, Ramin SM, Casey B et al. A multicenter, randomized trial of treatment for mild gestational diabetes. N Engl J Med 2009; 361(14):1339-48.
11. Horvath K, Koch K, Jeitler K, Matyas E, Bender R, Bastian H et al. Effects of treatment in women with gestational diabetes mellitus: systematic review and meta-analysis. BMJ 2010; 340:c1395.
12. Alwan N, Tuffnell DJ, West J. Treatments for gestational diabetes. Cochrane Database Syst Rev 2009; (3):CD003395.
13. McElduff A, Goldring J, Gordon P, Wyndham L. A direct comparison of the measurement of a random plasma glucose and a post-50 g glucose load glucose, in the detection of gestational diabetes. Aust N Z J Obstet Gynaecol 1994; 34(1):28-30.
14. Mathai M, Thomas TJ, Kuruvila S, Jairaj P. Random plasma glucose and the glucose challenge test in pregnancy. Natl Med J India 1994; 7(4):160-2.
15. Naylor CD, Sermer M, Chen E, Farine D. Selective screening for gestational diabetes mellitus. Toronto Trihospital Gestational Diabetes Project Investigators. N Engl J Med 1997; 337(22):1591-6
16. NVOG richtlijn "Diabetes en zwangerschap". http://nvog-documenten.nl/index.php?pagina=/richtlijn/pagina.php&fSelectTG_62=75&fSelectedSub=62&fSelectedParent=75
17. O'Sullivan JB, and Mahan C. Criteria for oral glucose tolerance test in pregnancy, Diabetes 1964; 13:278-285.
18. American Diabetes Association. Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 2009; 32(Supp 1):S62-S67.
19. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and Classification of Diabetes Mellitus; Report of a WHO consultation. Diabet Med 1998; 15(7):539-53.

20. International Association of Diabetes and Pregnancy Study Groups Consensus Panel. Recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010; 33:676-682
21. Sacks DA, Abu-Fadil S, Karten GJ, Forsythe AB, Hackett JR. Screening for gestational diabetes with the one-hour 50-g glucose test. Obstet Gynecol 1987; 70:89-93.
22. Nanda S, Savvidou M, Syngelaki A, Akolekar R, Nicolaides KH. Prediction of gestational diabetes mellitus by maternal factors and biomarkers at 11 to 13 weeks. Prenat Diagn 2011; 31(2):135-41.
23. HAPO Study Cooperative Research Group. Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study: associations with maternal body mass index. Br J Obstet Gynaecol 2010; 117(5):575-84.
24. Dodd JM, Crowther CA, Robinson JS. Dietary and lifestyle interventions to limit weight gain during pregnancy for obese or overweight women: a systematic review. Acta Obstet Gynecol Scand 2008; 87:702-6.
25. Hatem M, Anthony F, Hogston P, Rowe DJ, Dennis KJ. Reference values for 75 g oral glucose tolerance test in pregnancy. BMJ 1988; 296: 676-678.
26. Fachnie JD, Whitehouse FW, McGrath Z. Vomiting during OGTT in third trimester of pregnancy. Diabetes Care; 11:818.
26. Ohno MS, Sparks TN, Cheng YW, Caughey AB. Treating mild gestational diabetes mellitus: a cost-effectiveness analysis. Am J Obstet Gynecol 2011; 205(3):282.e1-7.

Chapter 11

Samenvatting, algemene discussie en toepassingen

Samenvatting

Diabetes gravidarum is geassocieerd met een hogere kans op perinatale en maternale complicaties tijdens de zwangerschap en durante partu.[1-3] Het is waarschijnlijk dat er voor zowel moeder als kind ook lange termijn gevolgen van diabetes gravidarum zijn: Er bestaat een predispositie voor obesitas, het metabool syndroom en diabetes mellitus.[4-7] Om goede zorg te kunnen bewerkstellingen tijdens de zwangerschap en de bevalling, maar ook in de periode postpartum en ten aanzien van de lange termijn, is (tijdige) identificatie van vrouwen met diabetes gravidarum van belang. Alleen als er een adequate methode bestaat om diabetes gravidarum op te sporen, kan tijdig een behandeling kan worden gestart om de kans op complicaties te reduceren.[8] Het belang hiervan wordt benadrukt doordat door de wereldwijd groeiende epidemie van obesitas en diabetes mellitus type 2 het aantal vrouwen met diabetes gravidarum ook zal toenemen.

Vrouwen met diabetes gravidarum hebben meestal geen klachten of symptomen. Dit bemoeilijkt het opsporen van de aandoening. De afgelopen jaren zijn er verschillende testen en strategieën toegepast, om vrouwen met diabetes gravidarum op te sporen. Het is onduidelijk welke strategie het meest effectief is. In proefschrift hebben wij onderzocht wat de meest (kosten-) effectieve manier is om diabetes gravidarum op te sporen. In dit hoofdstuk zullen de resultaten van het onderzoek worden samengevat en in de discussie zullen de resultaten in perspectief worden geplaatst. Ook zullen er suggesties gedaan worden voor toekomstig onderzoek.

In 1968 hebben Wilson en Jungner criteria opgesteld ten aanzien van de selectie van ziektebeelden die geschikt zijn voor screening in het algemeen. In die tijd stelden zij vast dat het idee van vroege opsporing en behandeling van ziekte simpel lijkt in de basis, maar dat de weg naar succesvolle bewerkstelliging, ondanks dat deze soms eenvoudig lijkt, zeker niet gemakkelijk is.[9] Meer dan 40 jaar later geldt dit nog steeds voor screening op diabetes gravidarum. Tot voorkort werden de aard en de incidentie van complicaties geassocieerd met diabetes gravidarum niet eenduidig bewezen geacht. Ook was het niet bewezen dat behandeling van diabetes gravidarum de kans op complicaties zou reduceren. Dit had tot gevolg dat er lange tijd geen uniform beleid was ten aanzien van screening, diagnostiek en behandeling.

De resultaten van drie grote studies en twee systematische reviews van de literatuur hebben de laatste jaren meer duidelijkheid gebracht over de risico's geassocieerd met diabetes gravidarum, en over het gunstige effect van behandeling.[1-3,10-12] Omdat diabetes gravidarum veelal asymptomatisch verloopt, is de enige manier om de aandoening tijdig op te sporen door middel van screening. Hiertoe zijn tot nu toe verschillende testen en

strategieën gebruikt, zoals de random glucose meting, de 50-g glucose belasting test, de nuchtere glucose meting en selectie van vrouwen door middel van risicofactoren. Het is onduidelijk welk van deze strategieën de beste is.

Het doel van dit proefschrift was om de accuratesse en de kosten van een aantal screening strategieën waaronder zowel glucose testen als risico schatting op basis van patiënt kenmerken te evalueren, om zo één adequate en kosten effectieve strategie te realiseren om vrouwen met diabetes gravidarum op te sporen, en de kans op complicaties als gevolg van de aandoening te reduceren.

We hebben hiertoe met data van een prospectieve cohort studie de accuratesse van twee screening testen direct met elkaar vergeleken, afgezet tegen de referentiestandaard (orale glucose tolerantie test, OGTT). We hebben tevens een aantal systematische reviews van de internationale literatuur verricht, om zo de accuratesse van verschillende bloedglucose testen te kunnen schatten en vergelijken. Op het gebied van de toepassing van risicofactoren in het diagnostisch proces hebben we een bestaand risicostratificatie systeem extern gevalideerd. Tevens hebben we zelf een klinisch predictie model ontwikkeld bestaande uit patiënt karakteristieken, om zo het individuele risico op diabetes gravidarum per patiënt te kunnen schatten. Verder hebben we vanuit de bestaande literatuur een overzicht gegeven van de resultaten van een aantal grote studies over de maternale en perinatale risico's geassocieerd met diabetes gravidarum en het effect van behandeling, en hebben we een inventarisatie verricht onder Nederlandse gynaecologen en verloskundigen om het huidig toegepaste beleid ten aanzien van detectie en behandeling van diabetes gravidarum in Nederland te evalueren. Tot slot hebben we een kosteneffectiviteit analyse verricht, waarbij we kosten van de verschillende testen en strategieën per voorkomen complicatie hebben vergeleken.

Hoofdstuk 1 geeft een overzicht van de indeling, en beschrijft het doel van dit proefschrift

In **hoofdstuk 2** worden de resultaten gepresenteerd van een studie over de directe vergelijking van twee screening testen voor diabetes gravidarum; de random glucose meting en de 50-g glucose belasting test. Dit onderzoek werd verricht met data van een prospectieve cohort studie waarin 1301 vrouwen werden geïncludeerd. Alle vrouwen ondergingen beide screening testen bij een amenorroeduur van 24 tot 28 weken. De referentiestandaard om diabetes gravidarum vast te stellen of uit te sluiten was de 75-g OGTT. De oppervlakte onder de receiver operating characteristic (ROC) curve was 0.88 (95% BI 0.83 - 0.93) voor de 50-g glucose belasting test, en 0.69 (95% BI 0.61 - 0.78) voor de random glucose meting. Het verschil tussen de beide oppervlakten onder de curve was 0.19 (95% BI 0.11 - 0.27) in het voordeel van de 50-g glucose belasting test. Er zijn twee andere studies die deze twee screening testen direct hebben vergeleken.

McElduff *et al*. vonden dat de 50-g glucose belasting test betere resultaten gaf in het opsporen van diabetes gravidarum.[13] Mathai *et al*. vonden juist, dat de oppervlakte onder de ROC curve groter was voor de random glucose meting indien vergeleken met de 50-g glucose belasting test bij 26 tot 30 weken amenorroeduur.[14] In beide studies was de referentiestandaard om diabetes gravidarum vast te stellen de 100-g OGTT. Gebaseerd op de resultaten van onze studie concluderen wij dat voor het opsporen van diabetes gravidarum de 50-g glucose belasting test geschikter is dan de random glucose meting.

In **hoofdstuk 3, 4 en 5** beschrijven we de resultaten van een aantal systematische reviews van de literatuur over verschillende screening testen voor diabetes gravidarum. Indien mogelijk hebben we bivariate meta-analyse verricht om samenvattende schattingen te berekenen van de sensitiviteit en specificiteit van de testen.

Hoofdstuk 3 behandelt de resultaten van een systematische review van de literatuur over de accuratesse van de random glucose meting als screening test voor diabetes gravidarum. We includeerden zes originele studies (totaal 3537 patiënten). Door het kleine aantal studies en de klinische heterogeniteit van de verschillende studies was het niet mogelijk om samenvattende schattingen van de accuratesse (sensitiviteit en specificiteit) te berekenen. De gerapporteerde accuratessematen in de individuele studies varieerden sterk. Voor 100% sensitiviteit, werd een specificiteit gerapporteerd van rond de 40%. Voor een sensitiviteit van 60% was de specificiteit maximaal 80%. Voor een hogere specificiteit (rond de 100%) was de sensitiviteit maximaal 20 tot 30%. Hoewel het aantal studies klein is en de geïncludeerde studies aanzienlijke klinische heterogeniteit vertonen, concluderen we dat één enkele random glucose meting niet geschikt is als screening test voor diabetes gravidarum.

Hoofdstuk 4 presenteert de resultaten van een systematische review en bivariate meta-analyse van de accuratesse van de 50-g glucose belasting test als screening test voor diabetes gravidarum. Er werden 26 studies geïncludeerd (totaal 13 564 vrouwen) die de 50-g glucose belasting test vergeleken met de 75- of 100-g OGTT als referentietest, voor 32 weken amenorroeduur. In de studies die alleen vrouwen includeerden met risicofactoren voor diabetes gravidarum was de gepoolde schatting van de sensitiviteit 0.74 (95% BI 0.62 - 0.87) en de gepoolde schatting van de specificiteit 0.77 (95% BI 0.66 - 0.89) voor een afkapwaarde van de 50-g glucose belasting test van 7.8 mmol/l. Bijbehorende likelihood ratios (LR) voor positief dan wel negatief test resultaat waren respectievelijk 3.2 (95% BI 2.0 - 5.2) and 0.34 (95% BI 0.22 - 0.53). In studies met inclusie van alle vrouwen was de gepoolde schatting van sensitiviteit 0.74 (95% BI 0.62 - 0.87) en van de specificiteit 0.85 (95% BI 0.80 - 0.91). Bijbehorende LR voor een positief dan wel negatief testresultaat waren respectievelijk 4.9 (95% BI 3.5 - 7.0)

en 0.31 (95% BI 0.20 - 0.47). Hoewel hogere accuratesse de voorkeur zou hebben, lijkt een detectie graad van 74% acceptabel. Om de 50-g glucose belasting test als definitieve diagnostische test toe te passen voor diabetes gravidarum (vervangen van de OGTT) zijn hogere accuratesse maten noodzakelijk. We concluderen dat de 50-g glucose belasting test geschikt is als screening test voor diabetes gravidarum, echter niet om de OGTT te vervangen.

In **hoofdstuk 5** geven we een overzicht van de resultaten van een systematische review en bivariate meta-analyse van de literatuur over de nuchtere glucose meting als screening test voor diabetes gravidarum. Er werden 16 studies geïncludeerd waarin de nuchtere glucose meting werd vergeleken met de referentiestandaard (de 75-g of 100-g OGTT) om diabetes gravidarum vast te stellen of uit te sluiten, voor een amenorroeduur van 32 weken. Totaal werden 25 560 vrouwen in de individuele studies geïncludeerd. Er was geen associatie tussen studie populatie (inclusie van vrouwen met risicofactoren of inclusie van alle vrouwen), afkapwaarde van de OGTT en de samenvattende schattingen van de sensitiviteit en de specificiteit van de nuchtere glucose meting. De gepoolde schattingen van sensitiviteit met een bivariaat regressiemodel waren 0.30 (95% BI 0.09 - 0.65), 0.75 (95% BI 0.60 - 0.86) en 0.92 (95% BI 0.81 - 0.97) voor afkapwaarden van de nuchtere glucose meting van respectievelijk > 5.0 mmol/L, 4.6-5.0 mmol/L en < 4.6 mmol/L. De gepoolde schattingen van specificiteit voor deze afkapwaarden waren respectievelijk 0.96 (95% BI 0.90 - 0.98), 0.70 (95% BI 0.47 - 0.86) en 0.45 (95% BI 0.27 - 0.65). Een geschikte screening test zou bij voorkeur een hoge sensitiviteit moeten hebben, maar niet ten koste van een te lage specificiteit, omdat een te lage specificiteit zorgt dat veel vrouwen een (overbodige) OGTT moeten ondergaan. Dit kan angst, ongemak en stress met zich mee brengen en is daarom ongewenst. We concluderen dat de accuratesse van de nuchtere glucose meting niet voldoende is om de OGTT te vervangen als enige, diagnostische test voor diabetes gravidarum. Toekomstig onderzoek zou zich kunnen richten op de waarde van de nuchtere glucose meting in specifieke subgroepen. Tevens zouden de mogelijkheden van de nuchtere glucose meting en de 50-g glucose belasting test in combinatie met andere screening strategieën kunnen worden onderzocht.

In **hoofdstuk 6** beschrijven de resultaten van een validatie studie. Met data van een prospectieve cohort studie hebben we een bestaand klinisch risicostratificatie model voor het schatten van het risico op diabetes gravidarum gevalideerd. Het risicostratificatie model was oorspronkelijk ontwikkeld in Canada.[15] In het kader van het risicostratificatie model werd aan alle zwangere vrouwen een risico score toegekend op basis van leeftijd, BMI en etniciteit. De capaciteit van het model werd getest in termen van discriminatie (onderscheidend vermogen) en calibratie (overeenkomst tussen klinische score en de geobserveerde kans op diabetes gravidarum). Het score systeem discrimineerde matig (oppervlakte onder de ROC curve 0.64 (95% BI 0.56 - 0.72)) en de calibratie was

beperkt. De screening strategie die was gebaseerd op het klinische risicostratificatie systeem reduceerde echter het aantal vrouwen dat een OGTT zou moeten ondergaan met 25% ten opzichte van universele screening, voor een vergelijkbare detectiegraad. Wij concluderen dat ondanks een matige discriminatie en beperkte calibratie van het klinische risicostratificatie systeem, de strategie gebaseerd op het systeem klinisch nuttig en toepasbaar lijkt. Een beter predictie model om de kans op diabetes gravidarum te schatten is wenselijk.

Hoofdstuk 7 beschrijft de ontwikkeling van een klinisch predictie model voor diabetes gravidarum. Het predictie model werd gegenereerd met data van een prospectieve cohort studie. We evalueerden de voorspellende waarde van het predictie model door middel van discriminatie en calibratie. Tevens evalueerden we de klinische consequenties van screening met het predictie model. Het risico op diabetes gravidarum kon worden geschat op basis van de volgende factoren: etniciteit, familie anamnese, diabetes gravidarum in de voorgeschiedenis, en BMI. Het predictie model had een oppervlakte onder de ROC curve van 0.77 (95% BI 0.69 - 0.85). De calibratie was goed (Hosmer en Lemeshow test, $p = 0.25$). Indien een OGTT zou worden verricht bij alle vrouwen met een kans op diabetes gravidarum van 2% of hoger, zou 43% van alle vrouwen in het cohort een OGTT moeten ondergaan en zou 75% van alle vrouwen met diabetes gravidarum worden opgespoord. We stellen vast dat het klinische predictie model een geschikte methode lijkt om vrouwen met een verhoogd risico op diabetes gravidarum op te sporen, en dat het model de potentie lijkt te hebben om de efficiëntie van selectie voor screening te verbeteren. Echter voordat een klinisch predictie model in de praktijk kan worden gebruikt zal het model extern gevalideerd moeten worden.

Hoofdstuk 8 geeft een overzicht van de literatuur over de kans op complicaties als gevolg van hyperglycemie in de zwangerschap, en het effect van behandeling. De resultaten van drie grote studies op dit gebied worden besproken. Eén cohort studie waarin meer dat 23 000 vrouwen zijn geïncludeerd laat zien dat er een lineair verband bestaat tussen de waarden van de OGTT en het risico op maternale en perinatale complicaties. Twee andere studies laten zien dat behandeling van (milde) diabetes gravidarum het risico op complicaties verkleint. In dit hoofdstuk bespreken we ook de resultaten van een survey die werd verricht onder gynaecologen en verloskundigen om de klinische praktijk ten aanzien van screening, diagnostiek en behandeling in Nederland te inventariseren. Dit onderzoek werd verricht voor de publicatie van de hernieuwde Nederlandse richtlijn "Diabetes en zwangerschap" in 2010.[16] De meerderheid van de gynaecologen en verloskundigen rapporteerden dat zij screening op diabetes gravidarum verrichten in het eerste en tweede trimester van de zwangerschap. Er bleek een grote variatie in het gebruik van testen en screening strategieën. Dit kwam overeen met onderzoeken die hiernaar zijn verricht in andere landen. De test die het meest werd gebruikt als screening

test was de random glucose meting. De test die het meest gebruikt werd als test om diabetes gravidarum vast te stellen was een ontbijt of een lunch test (43%), gevolgd door de 75-g OGTT (31%). We concluderen dat voor publicatie van de richtlijn "Diabetes en zwangerschap" in 2010 suboptimale testen en strategieën werden toegepast om diabetes gravidarum op te sporen. We hebben geen data over het gebruik van de verschillende testen na de publicatie van de nieuwe richtlijn.

Hoofdstuk 9 beschrijft de resultaten van een kosteneffectiviteitsanalyse. In dit hoofdstuk evalueerden we met behulp van een model de kosten en effecten van een aantal verschillende testen en strategieën, om zo te berekenen welke strategie het meest kosten effectief is om het aantal ernstige perinatale complicaties als gevolg van diabetes gravidarum te reduceren. De screening strategie gebaseerd op het klinische predictie model in combinatie met een nuchtere glucose meting en een OGTT indien de nuchtere waarde afwijkend is, was de strategie met de laagste kosten per voorkomen ernstige perinatale complicatie (samengestelde uitkomst van perinatale sterfte, schouderdystocie en geboorte trauma) (€26 172 per voorkomen ernstige complicatie). Een hoger aantal ernstige complicaties kan worden voorkomen indien universele screening met een OGTT wordt toegepast. De kosten per voorkomen ernstige complicatie zijn dan echter wel hoger (€43 171 per voorkomen ernstige perinatale complicatie). De kosteneffectiviteit hangt af van de "willingness to pay" per voorkomen ernstige perinatale complicatie. Indien schouderdystocie niet werd meegenomen als ernstige perinatale complicatie, was de strategie van het predictie model in combinatie met de nuchtere glucose meting nog steeds geassocieerd met de laagste kosten per voorkomen ernstige complicatie. Echter de kosten per voorkomen ernstige perinatale complicatie stegen voor alle strategieën (€65 430 per voorkomen ernstige perinatale complicatie bij de strategie waarin het predictie model wordt gecombineerd met een nuchtere glucose meting).

Discussie

De originele criteria voor diabetes gravidarum werden in 1964 opgesteld door O'Sullivan and Mahan, en waren gebaseerd op de 100-g OGTT.[17] Deze criteria selecteerden vrouwen die een risico hadden op diabetes type 2 later in hun leven, en waren niet gericht op het risico op perinatale en maternale complicaties in de zwangerschap en durante partu. Door de jaren heen hebben de American Diabetes Association (ADA) en de World Health Organisation (WHO) respectievelijk de 100-g OGTT en de 75-g OGTT aangenomen als referentie test voor diabetes gravidarum, beide met hun eigen criteria / afkapwaarden.[18,19] In 2010 heeft de International Association of the Diabetes in Pregnancy Study Group (IADPSG) nieuwe diagnostische criteria voor diabetes gravidarum voorgesteld. Deze criteria zijn gebaseerd op een 75-g OGTT.[20] Omdat de

studies beschreven in dit proefschrift verricht zijn vanaf 2006, hebben wij veelal gebruik gemaakt van de WHO criteria voor diabetes gravidarum, en in mindere mate van de ADA criteria. Om de resultaten van onze onderzoeken in het perspectief te zien van de recentelijk voorgestelde IADPSG criteria, zullen we hier de achtergrond en consequenties van deze criteria bespreken.

De IADPSG criteria zijn voorgesteld na uitvoerige analyses van de Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) studie.[20] In de HAPO studie werden meer dan 23 000 vrouwen geïncludeerd die tussen 24 en 28 weken zwangerschap een 75-g OGTT ondergingen. Het doel van de studie was om te de associatie te evalueren tussen maternale glucose waarden en de kans op nadelige uitkomsten van de zwangerschap. Er werd een continue associatie gevonden tussen maternale glucose waarden en het risico op nadelige zwangerschapsuitkomsten (o.a. macrosomie, sectio caesarea, klinische neonatale hypoglycemie, serum C-peptide > 90^{e} percentiel, pre-eclampsie, opname op de neonatale intensive care, en neonatale hyperbilirubinemie), ook bij waarden die niet vallen binnen de definitie van diabetes gravidarum.[2] De uitkomsten waren duidelijk, maar de klinische toepasbaarheid van de uitkomsten was lastig, omdat er meerdere maternale en perinatale uitkomsten zijn meegenomen, en er door de continue lineaire relatie geen evidente afkapwaarde kon worden vastgesteld voor het verhoogde risico.

Na uitvoerig internationaal overleg kwamen er twee aanbevelingen vanuit de IADPSG ten aanzien van screening op diabetes gravidarum. Het advies luidde dat alle vrouwen met risico factoren voor diabetes gravidarum vroeg in de zwangerschap gescreend moeten worden om zo niet eerder ontdekte diabetes type 2 op te sporen, omdat deze vrouwen een verhoogd risico lopen op een foetus met congenitale afwijkingen.[20] De tweede aanbeveling was dat alle vrouwen bij 24 tot 28 weken amenorroeduur een 75-g OGTT zouden moeten ondergaan. Indien één van de volgende waarden wordt overschreden is er sprake van diabetes gravidarum: nuchtere waarde >5.1 mmol/L; 1 uur waarde >10.0 mmol/L; 2 uur waarde >8.5 mmol/L. Deze afkapwaarden zijn vastgesteld op een odds ratio (OR) van 1.75 voor de primaire uitkomsten van de HAPO studie (geboortegewicht > 90^{e} percentiel, primaire sectio caesarea, klinische neonatale hypoglycemie, en serum C-peptide > 90^{e} percentiel).[20]

Hoewel de IADPSG consensus aanbevelingen een belangrijke stap zijn in het bereiken van een uniforme, evidence-based manier van screening en diagnostiek naar diabetes gravidarum, zijn de aanbevelingen internationaal nog niet unaniem geaccepteerd. Het belangrijkste punt van kritiek is dat indien alle vrouwen een OGTT ondergaan, en er slechts één van de drie metingen afwijkend hoeft te zijn om diabetes gravidarum vast te stellen, het aantal vrouwen met diabetes gravidarum door de nieuwe criteria substantieel zal toenemen (tot 17.8%) vergeleken met de WHO en de ADA criteria. Hierdoor zullen

de kosten geassocieerd met de screening en de behandeling enorm stijgen. Het effect van behandeling van diabetes gravidarum gediagnosticeerd met de nieuwe IADPSG criteria is nog niet onderzocht. Hoewel het effect van behandeling eerder is vastgesteld in twee gerandomiseerde onderzoeken, berusten deze onderzoeken op de WHO en de ADA criteria. Er kan niet worden vastgesteld dat dit effect zonder meer bevestigd zal worden indien de nieuwe IADSPG criteria worden ingesteld.[1,2,10-12]

Het belangrijkste verschil met de ADA criteria is dat de nieuwe IADPSG criteria slechts één in plaats van twee van de drie afwijkende waarden behoeft voor de diagnose diabetes gravidarum.[18] Het belangrijkste verschil met de WHO criteria is dat bij de WHO criteria de 2 uur afkapwaarde is vastgesteld op 7.8 mmol/L.[19] Hoewel de WHO criteria en de IADPSG wel degelijk anders zijn, heeft een recente meta-analyse laten zien dat zowel studies die de WHO criteria gebruiken als studies die de IADPSG criteria gebruiken een verhoogd risico op complicaties bij diabetes gravidarum laten zien, met vergelijkbare uitkomstmaten.[3]

Indien de IADPSG criteria wereldwijd geaccepteerd zullen worden, betekent dit een substantiële toename van het aantal OGTT's, met als gevolg dat de kosten zullen stijgen. Een manier om het aantal vrouwen dat een OGTT ondergaat (en de hiermee gepaarde gaande kosten) te beperken zou kunnen zijn door een selectie toe te passen op basis van risicofactoren en patiënt karakteristieken. Deze manier van aanpak zou bijvoorbeeld toepasbaar kunnen zijn in populaties waar de prevalentie van diabetes gravidarum laag is. Indien het risico op diabetes gravidarum nauwkeurig voorspeld kan worden, kunnen vrouwen met een hoog risico getest worden met een OGTT, terwijl vrouwen met een laag risico geen test hoeven te ondergaan. Dit vermindert de belasting voor de zwangere vrouwen, maar ook de kosten voor de maatschappij.

Een aantal studies heeft het gebruik van risicofactoren in de screening op diabetes gravidarum onderzocht. Slechts een klein aantal heeft de resultaten weergegeven in de vorm van predictie model of score systeem, waarin alle factoren gecombineerd worden en elke factor de juiste waarde toebedeeld krijgt.[15,21] Wij hebben een predictie model ontwikkeld door middel van multipele regressie analyse, waarmee het risico op diabetes gravidarum voor de individuele patiënt geschat kan worden op basis van patiënt karakteristieken. Het risico in ons cohort kon voorspeld worden op basis van etniciteit, familie anamnese, diabetes gravidarum in de voorgeschiedenis en BMI. Een beperking van onze studie lag in het feit dat in de originele studie niet bij iedereen standaard een diagnostische test werd verricht. De kans dat een diagnostische OGTT werd verricht was afhankelijk van de uitslagen van de twee screening testen die werden verricht. Hierdoor ontstond verificatie bias. Hoewel we gecorrigeerd hebben voor deze vorm van bias door middel van een multipele imputatie procedure, weten we niet in hoeverre dit de

resultaten van het predictie model heeft beïnvloed. De geschatte detectie graad van het predictie model was 75%. Dit is een hogere detectiegraad dan werd gevonden in studies die waarde van één individuele risicofactor als separate screening test hebben onderzocht. De oppervlakte onder de ROC curve was 0.77 (95% BI 0.69 - 0.85). Nanda *et al*. ontwikkelde een vergelijkbaar model om het risico op diabetes gravidarum te schatten. Dit model leverde vergelijkbare resultaten op (oppervlakte onder de ROC curve 0.79 (95% BI 0.76 - 0.82)). Zij stelden vast dat het risico op diabetes gravidarum geschat kon worden op basis van maternale leeftijd, BMI, etniciteit, diabetes gravidarum in de voorgeschiedenis of eerdere macrosomie, waarbij diabetes gravidarum in de voorgeschiedenis de sterkste voorspeller was.[22] Predictie modellen dienen prospectief gevalideerd te worden voordat zij toegepast kunnen worden in de klinische praktijk. De modellen moeten worden getest in populaties die verschillen van de populatie waarin het model ontwikkeld is, om zo de klinische toepasbaarheid in andere situaties te onderzoeken. Indien de IADPSG criteria wereldwijd geaccepteerd worden, zullen validatie studies bij voorkeur plaats moeten vinden in combinatie met deze criteria. Het belang van externe validatie werd bevestigd in onze eigen validatie studie (hoofdstuk 6). Hier hebben we een bestaand score systeem voor diabetes gravidarum gevalideerd, waarbij we vaststelden dat discriminatie en calibratie in onze populatie beperkt waren, zelfs al maakten twee van de drie factoren die wij later in ons eigen predictie model includeerden deel uit van het risicostratificatie systeem.

Eén van de onafhankelijke risicofactoren voor diabetes gravidarum is obesitas. Er gaan stemmen op dat obesitas niet zozeer een risicofactor is voor diabetes gravidarum, maar direct invloed heeft op de kans op een nadelige zwangerschapsuitkomst. Uit primaire en secundaire analyses van de HAPO studie blijkt echter dat zowel diabetes gravidarum als obesitas onafhankelijke risicofactoren zijn voor een nadelige zwangerschapsuitkomst. Hogere maternale BMI is geassocieerd met foetale groei en adipositas en pre-eclampsie.[23] Omdat interventie strategieën voor de behandeling van obesitas in de zwangerschap weinig succesvol zijn gebleken, is het van belang dat diabetes gravidarum opgespoord en behandeld wordt, zodat de kans op complicaties wordt verminderd.[24]

Uitkomsten van screening testen alsmede de resultaten van de diagnostische OGTT worden vaak geclassificeerd als "normaal" of "abnormaal", afhankelijk van de vooraf bepaalde afkapwaarde. De associatie tussen de glucose waarde en het risico op complicaties is echter continu.[2] Door deze continue, lineaire associatie is er geen evidente afkapwaarde voor het risico op complicaties aan te wijzen.[2] Dit is de reden waarom het niet onverdeeld duidelijk is bij welke mate van hyperglycemie behandeling plaats zou moeten vinden. Interventie studies waarbij vrouwen met diabetes gravidarum volgens de WHO en de ADA criteria werden behandeld lieten een reductie zien van de kans op complicaties.[1,10] Resultaten van een systematische review laten zien, dat

interventies ook effectief zijn bij vrouwen met glucose concentraties die onder de diagnostische waarde voor diabetes gravidarum liggen. In deze groep werd het aantal macrosome babies gereduceerd door behandeling van de hyperglycemie. Dit geeft het continue karakter van diabetes gravidarum aan. Het dichotomiseren van de resultaten van de testen lijkt alleen nodig om te beslissen welke vrouwen behandeld zouden moeten worden. Wellicht zou het continue karakter van risico schatting het continuüm van risico op complicaties beter weer kunnen geven. De mogelijkheden van risicoschatting lijken de moeite waard om verder te exploreren door bestaande predictie modellen extern te valideren en te verbeteren, zo mogelijk gecombineerd met testen die beter getolereerd worden dan de OGTT om zo vrouwen te selecteren die baat hebben bij een behandeling van hyperglycemie, zonder dat daar perse het label "diabetes gravidarum" op geplakt hoeft te worden.[25,26]

We hebben een aantal systematische reviews verricht om de accuratesse van drie verschillende bloedglucose testen te evalueren. We hebben tevens de random glucose meting en de 50-g glucose belasting test direct met elkaar vergeleken. We concluderen daaruit dat de random glucose meting niet geschikt is als screening test voor diabetes gravidarum. De accuratesse van de 50-g glucose belasting test en de nuchtere glucose meting waren vergelijkbaar. Onze kosteneffectiviteitanalyse heeft laten zien dat screening op basis van een predictie model gecombineerd met een nuchtere glucose meting de strategie was met de laagste kosten om een ernstige perinatale complicatie te voorkomen (€26 172 per voorkomen ernstige complicatie), echter niet alle potentieel vermijdbare complicaties werden voorkomen met deze strategie. Indien een OGTT verricht zou worden bij alle vrouwen zouden meer complicaties voorkomen worden. De kosten per voorkomen complicatie liggen dan hoger, namelijk op €43 171 per voorkomen complicatie. De beste strategie is afhankelijk van de "willingness to pay" per voorkomen complicatie. Onze kosteneffectiviteitsanalyse was gebaseerd op de WHO criteria en we hebben verschillende screening strategieën met elkaar vergeleken. Ohno *et al* verrichtten een kosteneffectiviteitsanalyse waarin werd gekeken naar de kosten en het effect van behandeling of geen behandeling voor diabetes gravidarum gebaseerd op de IADPSG criteria.[22] De primaire uitkomstmaat waren de incrementele kosten per "quality-adjusted life year" (QALY). Zij concludeerden in deze studie dat de incrementele kosten per QALY $20 412 waren. Dit werd als kosten effectief beschouwd (onder de kosteneffectiviteit grens van $100 000/QALY).

In dit onderzoek hebben we niet geëvalueerd wat de voorkeur zou zijn van de zwangere vrouwen. De OGTT wordt in het algemeen als onplezierige test omschreven, terwijl compliance een belangrijke factor is, ook in een diagnostisch traject.[25,26]

Voordat de vernieuwde richtlijn "Diabetes en zwangerschap" van de Nederlandse Vereniging voor Obstetrie en Gynaecologie (NVOG) verscheen in 2010, werden verschillende testen en strategieën toegepast in de screening en diagnostiek naar diabetes gravidarum. De meest gebruikte test van de random glucose meting. De meest gebruikte diagnostische test was de ontbijt- of de lunchtest (43%). Gebaseerd op de resultaten beschreven in dit proefschrift zijn wij van mening dat de random glucose meting geen plaats heeft in het tweede trimester als screening test voor diabetes gravidarum. De diagnostische test is bij voorkeur de OGTT, en niet een ontbijt- of een lunchtest omdat voor deze laatste testen de relatie tussen de uikomsten en de kans op perinatale en maternale complicaties niet is onderzocht. Overeenkomstig wordt in de laatste richtlijn van de NVOG het gebruik van de random glucose in het tweede trimester niet meer aangeraden, en wordt het gebruik van de OGTT als diagnostische test wel geadviseerd.

Toepassingen

Door middel van screening kunnen vrouwen met (een hoog risico op) diabetes gravidarum worden geïdentificeerd. Deze vrouwen kunnen behandeld worden om zo de kans op complicaties te verminderen. Hiertoe is een accurate en kosten effectieve test of strategie nodig, die ervoor zorgt dat zoveel mogelijk vrouwen met diabetes gravidarum worden opgespoord terwijl zo min mogelijk vrouwen onnodig worden belast.

Er zijn nieuwe criteria voor diabetes gravidarum voorgesteld door de IADPSG. Indien deze criteria internationaal worden overgenomen zullen zij geëvalueerd moeten worden wat betreft het effect op zwangerschapsuitkomsten en de kosten voor de gezondheidszorg.[15] Het effect van behandeling zal geëvalueerd moeten worden, evenals het effect van screening strategieën.

Op basis van de resultaten van onze onderzoeken lijkt er een rol weggelegd voor risicofactoren in de screening op diabetes gravidarum. Het is echter van groot belang, dat deze rol nader wordt onderzocht door middel van validatie van het predictie model of ontwikkeling van nieuw predictie model, bij voorkeur in grote observationele studies.

Uit onze kosteneffectiviteitsanalyse blijkt dat het incorporeren van risicofactoren in een screening strategie wel de kosten van de screening beperkt, maar dat er relatief weinig complicaties als gevolg van diabetes gravidarum voorkomen worden met deze strategie. Daarom raden wij op dit moment aan, om voor de screening op diabetes gravidarum in het tweede trimester een nuchtere glucose meting of een OGTT bij alle vrouwen te verrichten. De mogelijkheden van de combinatie van een predictie model met een nuchtere glucose meting kunnen verder worden onderzocht.

Referenties

1. Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS; Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) Trial Group. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med 2005; 352(24):2477-86.
2. Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR et al. HAPO Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008; 358(19):1991-2002.
3. Wendland EM, Torloni MR, Falavigna M, Trujillo J, Dode MA, Campos MA et al. Gestational diabetes and pregnancy outcomes-a systematic review of the World Health Organization (WHO) and the International Association of Diabetes in Pregnancy Study Groups (IADPSG) diagnostic criteria. BMC Pregnancy and Childbirth 2012; 12:23
4. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. Diabetes Care 2002; 25(10):1862-8.
5. Yogev Y, Visser GH. Obesity, gestational diabetes and pregnancy outcome. Semin Fetal Neonatal Med 2009; 14:77-84.
6. Vohr BR, Boney CM. Gestational diabetes: the forerunner for the development of maternal and childhood obesity and metabolic syndrome? J Matern Fetal Neonatal Med 2008; 21:149-57.
7. Hillier TA, Pedula KL, Schmidt MM, Mullen JA, Charles MA, Pettitt DJ. Childhood obesity and metabolic imprinting: the ongoing effects of maternal hyperglycemia. Diabetes Care 2007; 30(9):2287-92.
8. Hunt KJ, Schuller KL. The increasing prevalence of diabetes in pregnancy. Obstet Gynecol Clin North Am 2007; 34(2):173-99.
9. Wilson JM, Jungner YG. Principles and practice of mass screening for disease. Bol Oficina Sanit Panam 1968; 65(4):281-393.
10. Landon MB, Spong CY, Thom E, Carpenter MW, Ramin SM, Casey B et al. A multicenter, randomized trial of treatment for mild gestational diabetes. N Engl J Med 2009; 361(14):1339-48.
11. Horvath K, Koch K, Jeitler K, Matyas E, Bender R, Bastian H et al. Effects of treatment in women with gestational diabetes mellitus: systematic review and meta-analysis. BMJ 2010; 340:c1395.
12. Alwan N, Tuffnell DJ, West J. Treatments for gestational diabetes. Cochrane Database Syst Rev 2009; (3):CD003395.
13. McElduff A, Goldring J, Gordon P, Wyndham L. A direct comparison of the measurement of a random plasma glucose and a post-50 g glucose load glucose, in the detection of gestational diabetes. Aust N Z J Obstet Gynaecol 1994; 34(1):28-30.
14. Mathai M, Thomas TJ, Kuruvila S, Jairaj P. Random plasma glucose and the glucose challenge test in pregnancy. Natl Med J India 1994; 7(4):160-2.
15. Naylor CD, Sermer M, Chen E, Farine D. Selective screening for gestational diabetes mellitus. Toronto Trihospital Gestational Diabetes Project Investigators. N Engl J Med 1997; 337(22):1591-6
16. NVOG richtlijn "Diabetes en zwangerschap". http://nvog-documenten.nl/index.php?pagina=/richtlijn/pagina.php&fSelectTG_62=75&fSelectedSub=62&fSelectedParent=75
17. O'Sullivan JB, and Mahan C. Criteria for oral glucose tolerance test in pregnancy, Diabetes 1964; 13:278-285.
18. American Diabetes Association. Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 2009; 32(Supp 1):S62-S67.
19. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and Classification of Diabetes Mellitus; Report of a WHO consultation. Diabet Med 1998; 15(7):539-53.

20. International Association of Diabetes and Pregnancy Study Groups Consensus Panel. Recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010; 33:676-682
21. Sacks DA, Abu-Fadil S, Karten GJ, Forsythe AB, Hackett JR. Screening for gestational diabetes with the one-hour 50-g glucose test. Obstet Gynecol 1987; 70:89-93.
22. Nanda S, Savvidou M, Syngelaki A, Akolekar R, Nicolaides KH. Prediction of gestational diabetes mellitus by maternal factors and biomarkers at 11 to 13 weeks. Prenat Diagn 2011; 31(2):135-41.
23. HAPO Study Cooperative Research Group. Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study: associations with maternal body mass index. Br J Obstet Gynaecol 2010; 117(5):575-84.
24. Dodd JM, Crowther CA, Robinson JS. Dietary and lifestyle interventions to limit weight gain during pregnancy for obese or overweight women: a systematic review. Acta Obstet Gynecol Scand 2008; 87:702-6.
25. Hatem M, Anthony F, Hogston P, Rowe DJ, Dennis KJ. Reference values for 75 g oral glucose tolerance test in pregnancy. BMJ 1988; 296: 676-678.
26. Fachnie JD, Whitehouse FW, McGrath Z. Vomiting during OGTT in third trimester of pregnancy. Diabetes Care; 11:818.
27. Ohno MS, Sparks TN, Cheng YW, Caughey AB. Treating mild gestational diabetes mellitus: a cost-effectiveness analysis. Am J Obstet Gynecol 2011; 205(3):282.e1-7.

List of Authors and affiliations

List of Authors and affiliations

E. van Ballegooie,* Department of Internal Medicine, Bethesda Hospital, Hoogeveen

H.G. ter Brugge, Department of Obstetrics & Gynaecology, Isala Clinics, Zwolle

I.M. Evers, Department of Obstetrics & Gynaecology, Meander Medical Center, Amersfoort

J. Limpens, Medical Library & Dutch Cochrane Centre, Academic Medical Center, Amsterdam

M.D. Louwerse, Department of Obstetrics & Gynaecology, VU Medical Center, Amsterdam

H. van Meir, Department of Obstetrics & Gynaecology, Leiden University Medical Center, Leiden

B.W.J. Mol, Department of Obstetrics & Gynaecology, Academic Medical Center, Amsterdam

B.C. Opmeer, Clinical Research Unit, Academic Medical Center, Amsterdam

S.M. Prins, Faculty of Medicine, University of Amsterdam, Amsterdam

J.B. Reitsma, Julius Center, University Medical Center Utrecht, Utrecht

M.J.M. Serlie, Department of Endocrinology & Metabolism. Academic Medical Center, Amsterdam

H.W. de Valk, Department of Internal Medicine. University Medical Center Utrecht, Utrecht

S.M.C. Vijgen, Dutch Healthcare Insurance Board (CVZ), Diemen.

G.H.A. Visser, Department of Obstetrics & Gynaecology. University Medical Center Utrecht, Utrecht

Y Yilmaz, Faculty of Medicine, University of Amsterdam, Amsterdam

E.J.K. Zweers,* Department of Internal Medicine, Bronovo Hospital, Den Haag

List of abbreviations

List of abbreviations

ADA	American Diabetes Association
AUC	Area under the curve
BMI	Body Mass Index
BWM	Ben Willem Mol
CI	Confidence Interval
DF	Degrees of freedom
FGT	Fasting glucose test
GCT	Glucose challenge test
GDM	Gestational Diabetes Melitus
HAPO	Hyperglycemia and Adverse Pregnancy Outcomes
HVM	Hélène van Meir
IADPSG	International Association of the Diabetes and Pregnancy Study Groups
ICER	Incremental cost-effectiveness ratio
IGT	Impaired glucose tolerance
JL	Jacqueline Limpens
LR	Likelihood ratio
MDL	Marjoleine Louwerse
mmol/L	millimole / liter
MVL	Marsha van Leeuwen
NICE	National Institute for Health and Clinical Excellence
NVOG	Nederlandse Vereniging voor Obstetrie en Gynaecologie
OGTT	Oral glucose tolerance test
OR	Odds Ratio
QALY	Quality-Adjusted Life Year
QUADAS	Quality Assessment tool for Diagnostic Accuracy Studies
RGT	Random glucose test
ROC	Receiver Operating Characteristic
SD	Standard deviation
SE	Standard error
SAS	Statistical Analysis System
SPSS	Statistical Package for the Social Sciences
T2D	Type 2 diabetes
WHO	World Health Organisation
WTP	Willingness to pay
YY	Yildirim Yilmaz

List of publications

List of publications

van Leeuwen M, Opmeer BC, Coppus SF, Mol BW. Fasting capillary glucose as a screening test for gestational diabetes mellitus *British Journal of Obstetrics and Gynaecology 2007; 114(3):372*.

van Leeuwen M, Zweers EJ, Opmeer BC, van Ballegooie E, ter Brugge HG, de Valk HW, Mol BW, Visser GH. Comparison of accuracy measures of two screening tests for gestational diabetes mellitus. *Diabetes Care 2007; 30(11):2779-84*.

van Leeuwen M, Opmeer BC, Zweers EJ, van Ballegooie E, ter Brugge HG, de Valk HW, Visser GH, Mol BW. External validation of a clinical scoring system for the risk of gestational diabetes mellitus. *Diabetes Research and Clinical Practice 2009; 85(1):96-101*.

van Leeuwen M, Opmeer BC, Zweers EJ, van Ballegooie E, ter Brugge HG, de Valk HW, Visser GH, Mol BW. Estimating the risk of gestational diabetes mellitus: a clinical prediction model based on patient characteristics and medical history. *British Journal of Obstetrics and Gynaecology 2010; 117(1):69-75*.

van Leeuwen M, Prins SM, de Valk HW, Evers IM, Visser GH, Mol BW. Gestational diabetes mellitus: treatment reduces the risk of complications. *Nederlands Tijdschrift voor Geneeskunde 2011; 155(9):A2291*.

van Leeuwen M, Opmeer BC, Yilmaz Y, Limpens J, Serlie MJ, Mol BW. Accuracy of the random glucose test as screening test for gestational diabetes mellitus. A systematic review. *European Journal of Obstetrics & Gynecology and Reproductive Biology 2011; 154(2):130-5*.

van Leeuwen M, Louwerse MD, Opmeer BC, Limpens J, Serlie MJ, Reitsma JB, Mol BW. Glucose challenge test for detecting gestational diabetes mellitus. A systematic review. *British Journal of Obstetrics and Gynaecology 2012; 119(4):393-401*.

Dankwoord

Dankwoord

Het boekje is klaar, wat een heerlijk idee! Dit proefschrift was er niet gekomen zonder de steun van velen. Ik wil een aantal mensen in het bijzonder bedanken.

Mijn promotor **Prof. dr. B.W.J. Mol**, Ben Willem. Zes jaar geleden nam je mij aan na een gesprek van 10 minuten. Dit bliksemgesprek was tekenend voor de snelheid waarmee jij denkt en handelt. Zelf was ik af en toe wat minder snel, maar nu is het proefschrift af! Ik wil je bedanken voor de inspirerende samenwerking. Jouw bezieling voor wetenschappelijk onderzoek is groot en daardoor zeer motiverend. Door jouw scherpe blik, doorzettingsvermogen en het voor ogen houden van de big-picture heb je het wetenschappelijk onderzoek binnen de verloskunde en gynaecologie nationaal en internationaal naar een hoog niveau gebracht. Ik ben er trots op dat ik daar deel van heb mogen uitmaken!

Mijn promotor **Prof. dr. G.H.A. Visser**, Gerard. Wat meer op de achtergrond ben jij zeer belangrijk geweest voor het welslagen van deze promotie. Naast het ter beschikking stellen van de data die hebben gediend als basis voor enkele hoofdstukken van dit proefschrift, heb ik veel gehad aan jouw enorme kennis over diabetes en zwangerschap. Je goede adviezen zorgden ervoor dat het klinisch epidemiologisch onderzoek ook daadwerkelijk klinisch toepasbaar bleef. Bedankt daarvoor.

Dr. B.C. Opmeer, Brent. Als co-promotor ben jij zeer nauw betrokken geweest bij het tot stand komen van dit proefschrift. Ik heb veel van je geleerd op het gebied van de klinische epidemiologie en statistiek. Ik ben je dankbaar voor je onmisbare hulp en geduld bij het analyseren van de data en het interpreteren van de analyses. Jouw snelle maar zeer punctuele correcties en adviezen in het laatste deel van het traject maakten het voor mij mogelijk om veel werk in een korte tijd te verzetten en zo het proefschrift af te maken. Heel veel dank daarvoor!

Dr. K.W.M. Bloemenkamp, Kitty, eigenlijk ben jij het die aan de basis staat van mijn loopbaan als onderzoeker. Door jouw enthousiaste begeleiding van mijn wetenschappelijke stage in 2005 kwam ik erachter dat onderzoek doen leuk kon zijn, ik had me geen betere en leukere stage kunnen wensen! Dankzij jou kwam ik in contact met Ben Willem, en kreeg ik de kans me verder te ontplooien binnen dit vakgebied. Ik ben je daar erg dankbaar voor!

De leden van de promotiecommissie: **Prof. dr. J.A.M van der Post, Prof. dr. H.P. Sauerwein, dr. I.M. Evers, dr. G. ter Riet en Prof. dr. A. Franx** wil ik bedanken voor het beoordelen van het manuscript en het plaatsnemen in de commissie. I would like to

thank **Professor B.E. Metzger** for crossing the Atlantic and being here today. You are one of the leading experts in the field, and it is a great honor that you are one of my opponents.

De mede auteurs van de verschillende artikelen wil ik bedanken voor hun bijdrage aan het onderzoek. Speciaal wil ik **Drs. E.J.K. Zweers*, Dr. E. van Ballegooie*, Drs. H.G. ter Brugge en dr. H.W. de Valk** bedanken voor hun input en het ter beschikking stellen van de data die de basis waren voor enkele hoofdstukken van dit proefschrift. Harold, ook bedankt voor je ideeën en hulp bij hoofdstuk 8. Verder **Dr. I.M. Evers**, Inge. Bedankt voor je bijdrage aan hoofdstuk 8 en 9, maar vooral voor je praktische kijk op dingen en voor de vrolijke noot in de soms zakelijke academische wereld!

Het **secretariaat AMC** en in het bijzonder **Marjan du Prie-Olthof** wil ik bedanken voor de hulp bij de voorbereidingen voor deze promotie. **Chris Bor** bedankt voor het verzorgen van de lay-out van het proefschrift. **Jana en Lex** heel veel dank voor het ontwerpen van de prachtige kaft! Het is een supermooi boek geworden!

De gynaecologen en andere medewerkers van de afdeling verloskunde en gynaecologie van het **Kennemer Gasthuis** wil ik bedanken voor de goede sfeer waarin ik het eerste deel van mijn opleiding heb gevolgd, het was een leerzame en zeker ook een leuke tijd. Ik kijk er naar uit om voor het laatste deel van mijn opleiding weer bij jullie terug te komen!

Stafleden, fellows, verloskundigen, verpleging, doktersassistenten en secretariaat van het **AMC** bedankt voor de leerzame opleidingstijd. Ik prijs me gelukkig dat ik elke dag met zoveel ambitieuze en stimulerende mensen kan werken.

Mijn **collega arts-assistenten gynaecologie** van het Kennemer en het AMC dank ik voor de mooie assistententijd die we samen hebben, zowel binnen het ziekenhuis als op borrels, feestjes en weekendjes weg. Ik ben ontzettend blij dat ik zulke fijne collega's heb!

Alle **onderzoekers van het AMC** (jullie zijn met veel teveel om op te noemen...) wil ik bedanken voor de mooie jaren! De hulp bij een statistisch probleem, de mogelijkheid om even te spuien (en vol goede moed weer verder te gaan), cursussen, congressen, borrels en gezelligheid waren vanzelfsprekend, maar van grote waarde! Dankzij jullie heb ik met plezier in dat veel te kleine hokje op H4 gezeten.

Mijn paranimfen **Arianne Lim** en **Menke Hazewinkel**. Lieve **Arianne**, kritische alleskunner! In april 2006 begonnen we als arts onderzoeker op heel verschillende projecten op de afdeling verloskunde. Het was heerlijk een veelal gelijkgestemd

iemand te hebben waarmee ik lief en leed kon delen op onderzoeksgebied, maar ook daarbuiten. Even heel hard zeiken over iets, om er vervolgens keihard om te kunnen lachen heeft bij mij de motivatie er altijd in weten te houden. Lieve Ari, je bent een uniek persoon. Intelligent, ambitieus en eigenwijs. Ik ben blij dat jij vandaag naast mij staat! Lieve **Menk**, talentvolle duizendpoot! Twee jaar lang zaten we samen op 2 m^2 met een kilometer aan mappen en twee loeiende computers. Ging super soepel, dat had ik met niemand anders volgehouden. Het was het fijn iemand (letterlijk) binnen handbereik te hebben die altijd klaar stond met waardevolle adviezen, wetenschappelijk of privé. Lieve Menk, ik bewonder je gedrevenheid en jouw talent voor het vak. Ik ben blij dat we nog dagelijks samenwerken in de kliniek en trots dat je mijn paranimf bent!

Er zijn veel mensen die niet betrokken zijn geweest bij mijn onderzoek, maar wel heel belangrijk zijn in mijn leven. Al mijn lieve **vrienden en vriendinnen**, jullie zijn er altijd! Voor leuke en soms ook minder leuke dingen. Jullie zorgen voor gezelligheid en ontspanning. Samen met jullie lachen, borrelen, dansen, eten en een goed gesprek op z'n tijd zorgt ervoor dat het leven leuk is. Ik prijs me gelukkig met zoveel lieve vrienden, dank daarvoor, ik zou jullie niet kunnen missen!

Lieve **Willem, Anneke en Maria**. Wat bof ik met zo'n lieve, nuchtere en betrokken schoonfamilie. Bij jullie is alles vanzelfsprekend en onvoorwaardelijk. Bedankt dat jullie er altijd voor ons zijn!

Lieve **Lennart en Ine**. Mijn lieve kleine broertje en schoonzus. Ook al zien we elkaar misschien minder vaak dan we zouden willen, de momenten samen zijn gezellig en vooral relaxed. Ik vind het super om te zien hoe jullie het leven samen leven! Lieve **Ine,** thanks voor het vertrouwen in de sneeuw! Lieve **Len,** denk jij ook nog wel eens aan al die vriendinnetjes die jij vroeger voor mij maakte op de camping? Ik heb inmiddels veel van jouw sociale vaardigheden afgekeken en dat komt me elke dag van pas!

Mijn lieve **ouders.** Door jullie onvoorwaardelijke steun en liefde heb ik de kans gehad om mezelf te ontwikkelen en te worden wie ik ben. Jullie hebben mij altijd het vertrouwen en de vrijheid gegeven om mijn eigen keuzes te maken, maar ook de zekerheid dat jullie er altijd voor me zijn. **Mam,** jij bent zonder twijfel de meest attente en zorgzame persoon die ik ooit zal leren kennen. **Pap,** jouw relativeringsvermogen is iets waar ik altijd op terug kan vallen, en je beroemde uitspraak "altijd rustig blijven" hoor ik regelmatig in mijn achterhoofd. Dank jullie wel voor het bijbrengen van heel veel goede dingen, en alles wat daardoor mogelijk is geworden!

Lieve **Norb,** mijn allerliefste schat van de wereld en onvoorwaardelijke steun en toeverlaat! Wat ben ik blij en zielsgelukkig dat wij samen zijn!

Curriculum Vitae

Marsha van Leeuwen was born on the 22nd of June 1979 in Den Haag. After graduation from high school in 1997 (Erasmus College, Zoetermeer), she studied Pharmacy at the University of Utrecht. In 1998 she started medical school at the Rijksuniversiteit Groningen. In 2005 she obtained her medical degree.

After travelling in Australia and New Zealand she started in 2006 as a PhD student at the Department of Obstetrics of the Academic Medical Center on a project entitled "Screening for Gestational Diabetes Mellitus". The project was supervised by Prof. dr. B.W.J. Mol (promotor), Prof. dr. G.H.A. Visser (promotor, UMCU) and Dr. B.C. Opmeer (co-promotor) and resulted in this thesis. After two years of full-time research, Marsha started working as a resident at the department of Obstetrics and Gynaecology of the Kennemer Gasthuis in Haarlem. She returned to the Academic Medical Center in January 2009 for another period of research.

In July 2009 she started her residency in Obstetrics and Gynaecology at the Kennemer Gasthuis in Haarlem (Head of the department dr. J.P. Lips). From October 2010 she continued her residency in the Academic Medical Center (Head of the department Prof. dr. M.J. Heineman). Marsha van Leeuwen lives with Norbert Fühler in Amsterdam.