

Comparison of Accuracy Measures of Two Screening Tests for Gestational Diabetes Mellitus

MARSHA VAN LEEUWEN, MD¹
 EGBERT J.K. ZWEERS, MD^{2,3}
 BRENT C. OPMEER, PHD⁴
 EVERT VAN BALLEGOOIE, MD, PHD^{5,6}

HENK G. TER BRUGGE, MD⁷
 HAROLD W. DE VALK, MD, PHD²
 BEN W.J. MOL, MD, PHD^{8,9}
 GERARD H.A. VISSER, MD, PHD⁸

OBJECTIVE — To compare the 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 preexisting diabetes in two perinatal centers 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 1 week. Furthermore, the OGTT was performed in a random sample of women in whom both screening tests were normal. GDM was considered present when the OGTT (reference test) exceeded predefined threshold values. Receiver operating characteristic (ROC) analysis was used to evaluate the performance of the two screening tests. The results were corrected for verification bias.

RESULTS — We included 1,301 women. The OGTT was performed in 322 women. After correction for verification bias, the random glucose test showed an area under the ROC curve of 0.69 (95% CI 0.61–0.78), whereas the glucose challenge test had an area under the curve of 0.88 (0.83–0.93). There was a significant difference in area under the curve of the two tests of 0.19 (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.

Diabetes Care 30:2779–2784, 2007

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; neonatal hypoglycemia; and jaundice. In addition, women with GDM are at substantially higher risk to develop diabe-

tes 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, obese, close relative with diabetes, history of GDM or 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 2-year study period were invited to participate. Women known to have preexisting diabetes were excluded from the study, as well as those who had not reported for prenatal care in one of two participating hospitals before 24 weeks of gestation. Only women who de-

From the ¹Department of Obstetrics and Gynaecology, Academic Medical Centre, Amsterdam, the Netherlands; the ²Department of Internal Medicine, University Medical Centre, Utrecht, the Netherlands; the ³Department of Internal Medicine, Bronovo Hospital, the Hague, the Netherlands; the ⁴Department of Clinical Epidemiology and Biostatistics, Academic Medical Centre, Amsterdam, the Netherlands; the ⁵Department of Internal Medicine, Isala Clinics, Zwolle, the Netherlands; the ⁶Department of Internal Medicine, Bethesda Hospital, Hoogeveen, the Netherlands; the ⁷Department of Obstetrics and Gynaecology, Isala Clinics, Zwolle, the Netherlands; the ⁸Department of Perinatology and Gynaecology, University Medical Centre, Utrecht, the Netherlands; and the ⁹Department of Obstetrics and Gynaecology, Máxima Medical Centre, Veldhoven, the Netherlands.

Address correspondence and reprint requests to Marsha van Leeuwen, MD, Academic Medical Centre, Department of Obstetrics and Gynaecology, Rm. H4-255, Meibergdreef 9, 1105 AZ Amsterdam, Netherlands. E-mail: marsha.vanleeuwen@amc.uva.nl.

Received for publication 22 March 2007 and accepted in revised form 31 July 2007.

Published ahead of print at <http://care.diabetesjournals.org> on 13 August 2007. DOI: 10.2337/dc07-0571.

Abbreviations: GDM, gestational diabetes mellitus; OGTT, oral glucose tolerance test; ROC, receiver operating characteristic.

A table elsewhere in this issue shows conventional and Système International (SI) units and conversion factors for many substances.

© 2007 by the American Diabetes Association.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

livered after 28 weeks of gestation were included in the analysis.

Data

At intake, the following information was obtained: obstetric history, family history of diabetes, ethnicity (categorized as Caucasian or non-Caucasian), height, self-reported weight (before pregnancy), age, and smoking habits (categorized as smoking or nonsmoking). 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 1 and 5 min, 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 ≥ 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 ≥ 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 1 h 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 2-h 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-h overnight fast and 3 days of minimal 150- to 200-g carbohydrate diet. Plasma glucose was determined before and 2 h 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 World Health Organization criteria (< 7.8 mmol/l after 12-h overnight fast and ≥ 7.8 mmol/l at 2 h after administration of a 75-g glucose-containing solution). These criteria were also applied in the random-

ized controlled trial in which treatment of GDM was considered beneficial (1). Venous plasma glucose concentration in all tests was evaluated via glucose oxidase method (Vitros; Ortho-Clinical Diagnostics, Amersham, U.K.) 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).

As 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. We performed the OGTT (reference test) 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 screening 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 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 means \pm SD. We constructed two-by-two tables for abnormal and normal test results on the random glucose test and the 50-g glucose screening 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% 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.

RESULTS

— We included 1,305 women. Four women were excluded from analysis because they delivered before 28 weeks of gestation. Data from 1,301 women were used for further analysis. Patient characteristics are presented in Table 1. Thereby, the distribution of patient characteristics within the classification groups of the reference test (OGTT) can be compared.

Figure 1 displays the flow of patients in our study based on the results of the subsequent diagnostic tests. Of all 1,301 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 1,169 (89.9%) and 1,295 (99.5%) of the 1,301 women, respectively. We used the results of the random glucose test obtained at intake for the 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 > 6.8 mmol/l. The 50-g oral glucose challenge test was performed in 1,281 women (98.5%).

There were 37 of 1,301 women (2.8%) who had an abnormal random glucose test, whereas 167 of 1,281 women (13.0%) had an abnormal 50-g glucose challenge test. There were 184 women (14.1%) with at least one abnormal 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 184 women (79.3%) with an abnormal screening test result and a subgroup of

Table 1—Demographics before correction for verification bias

	GDM present	GDM not present	GDM not verified	Total
<i>n</i>	46	276	979	1,301
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)	1,132 (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)	1,033 (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)	1,062 (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 spontaneous abortion	15 (32.6)	84 (30.4)	317 (32.4)	416 (32.0)
No previous spontaneous abortion	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* (%).

176 women with two negative screening tests (Fig. 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 value 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 nonrandom 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%). To obtain imputed data that are in line with the incidence of GDM in the Netherlands (es-

timated to be ~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 (>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 with the random glucose test (70.2% [95% CI 57.1–83.3] vs. 14.6% [4.6–24.6]). The random glucose test had less

false-positive test results and was therefore more specific (97.6% [96.6–98.5] vs. 89.1% [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 [0.83–0.93]) than for the random glucose test (0.69 [0.61–0.78]). There was a significant difference in the areas under the curve of the two tests of 0.19 (0.11–0.27).

CONCLUSIONS— 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

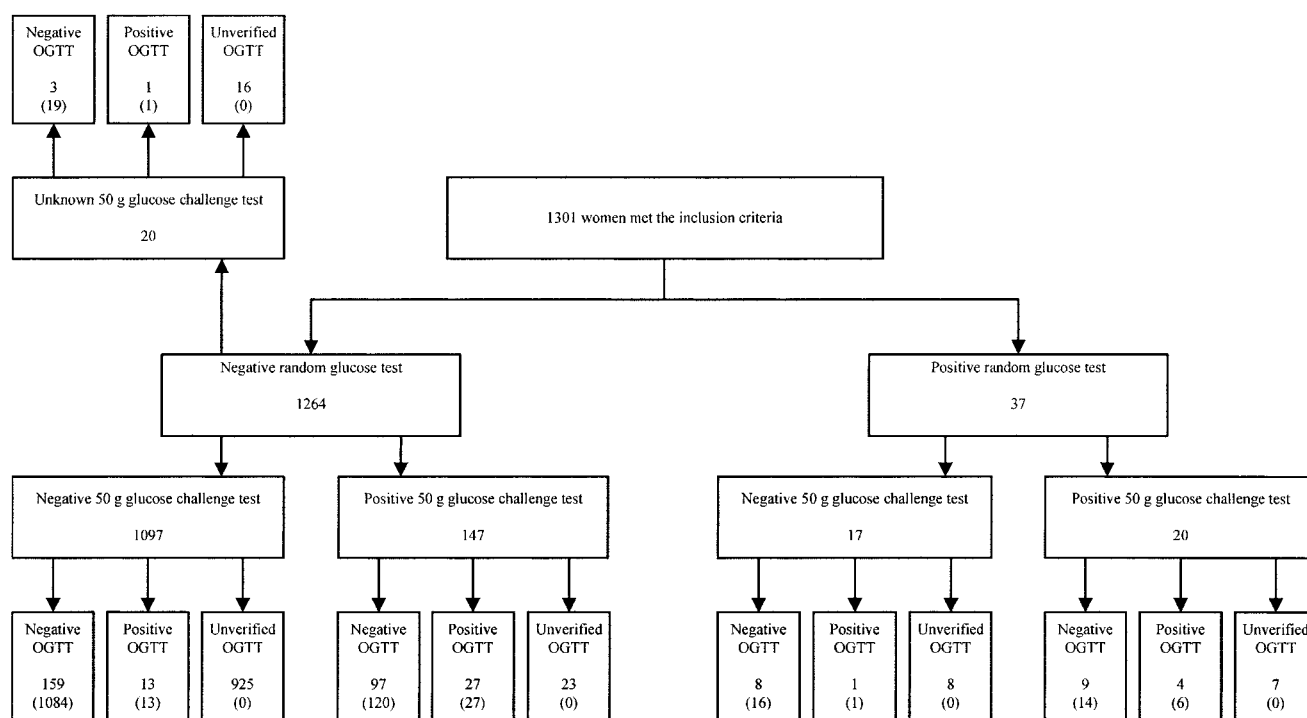


Figure 1—Screening and diagnostic test results before and after correction for verification bias. The figures in the diagram represent the number of women with the specific combination of test results before (after) correction of verification bias. Figures between parentheses represent the number of women after correction for verification bias.

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 the present study is the number of missing reference tests, due to which verification bias oc-

curred. 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 of diseased to nondiseased 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 or, in other words, if the chance of verification is truly independent of other factors such 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 at the individual level, accounting for factors that influence the chance of verification, imputation techniques can be used to

Table 2—Results of the 2 × 2 table and accuracy measures calculated after correction for verification bias

	Random glucose test				1-h 50-g glucose challenge test			
	OGTT abnormal	OGTT normal	OGTT unverified	Total	OGTT abnormal	OGTT normal	OGTT unverified	Total
Screening test abnormal	7	30	0	37	33	134	0	167
Screening test normal	41	1,223	0	1,264	14	1,100	0	1,114
Total	48	1,253	0	1,301	47	1,234	0	1,281
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 for an abnormal test result	6.1 (2.8–13.2)				6.5 (5.1–8.3)			
Likelihood ratio for a 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.

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, consequently implying 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 (>75 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 to 9% (1). In Western countries such as the Netherlands, the incidence is more often toward 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 incidence 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 with random glucose testing. To our knowledge, these two screening tests have only been equated in the same sample two times before. McElduff et al. (20) found their results in favor of the 50-g challenge test, whereas Mathai et al. (21) 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. A number of studies compared the 50-g glucose challenge test with measurement of fasting glucose. Perucchini et al. (22) found the results in favor of the fasting glucose measurement, whereas Rey et al. (23) showed the 50-g glucose challenge test to be superior. Other studies investigating the test characteristics of the glucose challenge test reported sensitivities ranging from 58 to 80% (24,25) for a specificity of ~65% (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 <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 that of the glucose challenge test. Nasrat et al. (26) 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 ≥ 2 h postprandial. Jowett et al. (27) also concluded that random glucose measurement is not sufficiently sensitive for screening on GDM. 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 because five of six women with GDM would still be missed.

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 the preferred screening test for GDM.

Acknowledgments— This research was supported by a grant from Novo Nordisk, Alphen aan den Rijn, the Netherlands, and by grant 917.46.346 in the VIDU Program of ZonMW, the Hague, the Netherlands. The funding sources had no involvement in the design, analysis, or reporting of this study.

Parts of this article were presented in abstract form at the 4th International Symposium of Diabetes and Pregnancy, which was held in Istanbul, Turkey, 29–31 March 2007.

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* 352:2477–2486, 2005
2. Esakoff TF, Cheng YW, Caughey AB: Screening for gestational diabetes: different cut-offs for different ethnicities? *Am J Obstet Gynecol* 193:1040–1044, 2005
3. Jensen DM, Damm P, Sorensen B, Molsted-

4. Pedersen L, Westergaard JG, Korsholm L, Ovesen P, Beck-Nielsen H: 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* 20:51–57, 2003
5. 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* 185:604–607, 2001
6. Tuffnell DJ, West J, Walkinshaw SA: Treatments for gestational diabetes and impaired glucose tolerance in pregnancy. *Cochrane Database Syst Rev* CD003395, 2003
7. 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* 101:380–392, 2003
8. Langer O, Yogev Y, Most O, Xenakis EM: Gestational diabetes: the consequences of not treating. *Am J Obstet Gynecol* 192:989–997, 2005
9. Naylor CD, Sermer M, Chen E, Farine D: Selective screening for gestational diabetes mellitus: Toronto Trihospital Gestational Diabetes Project Investigators. *N Engl J Med* 337:1591–1596, 1997
10. 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* 121:34–37, 2005
11. Santos-Ayazagoitia 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* 74:322–328, 2006
12. American Diabetes Association: Gestational diabetes mellitus. *Diabetes Care* 26 (Suppl. 1):S103–S105, 2003
13. Begg CB: Biases in the assessment of diagnostic tests. *Stat Med* 6:411–423, 1987
14. Rubin DB: *Multiple Imputation for Nonresponse in Surveys*. New York, John Wiley and Sons, 1987
15. Schafer JL: *Analysis of Incomplete Multivariate Data*. New York, Chapman and Hall, 1997
16. Frank Harrell: *Regression Modeling Strategies*. New York, Springer, 2001
17. Begg CB, Greenes RA: Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics* 39:207–215, 1983
18. Pepe MS: *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford, U.K., Oxford University Press, 2003
19. 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* 10:201–211, 1990

19. Pepe MS, Alonzo TA: Comparing disease screening tests when true disease status is ascertained only for screen positives. *Biostatistics* 2:249–260, 2001
20. 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* 34: 28–30, 1994
21. Mathai M, Thomas TJ, Kuruvila S, Jairaj P: Random plasma glucose and the glucose challenge test in pregnancy. *Natl Med J India* 7:160–162, 1994
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* 319:812–815, 1999
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* 37:780–784, 2004
24. de Sere day MS, Damiano MM, Gonzalez CD, Bennett PH: Diagnostic criteria for gestational diabetes in relation to pregnancy outcome. *J Diabetes Complications* 17:115–119, 2003
25. Maresh M: Screening for gestational diabetes mellitus. *Semin Fetal Neonatal Med* 10:317–323, 2005
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* 95:855–860, 1988
27. Jowett NI, Samanta AK, Burden AC: Screening for diabetes in pregnancy: is a random blood glucose enough? *Diabet Med* 4:160–163, 1987