Screening Tests for Gestational Diabetes: A Systematic Review for the U.S. Preventive Services Task Force

Lois Donovan, MD; Lisa Hartling, PhD; Melanie Muise, MA; Alyssa Guthrie, MSSc; Ben Vandermeer, MSc; and Donna M. Dryden, PhD

Background: A 50-g oral glucose challenge test (OGCT) is a widely accepted screening method for gestational diabetes mellitus (GDM), but other options are being considered.

Purpose: To systematically review the test characteristics of various screening methods for GDM across a range of recommended diagnostic glucose thresholds.

Data Sources: 15 electronic databases from 1995 to May 2012, reference lists, Web sites of relevant organizations, and gray literature.

Study Selection: Two reviewers independently identified English-language prospective studies that compared any screening test for GDM with any reference standard.

Data Extraction: One reviewer extracted and a second reviewer verified data from 51 cohort studies. Two reviewers independently assessed methodological quality.

Data Synthesis: The sensitivity, specificity, and positive and negative likelihood ratios for the OGCT at a threshold of 7.8 mmol/L [140 mg/dL] were 70% to 88%, 69% to 89%, 2.6 to 6.5, and 0.16 to 0.33, respectively. At a threshold of 7.2 mmol/L [130 mg/dL], the test characteristics were 88% to 99%, 66% to 77%, 2.7 to 4.2, and 0.02 to 0.14, respectively. For a fasting plasma glucose threshold of 4.7 mmol/L (85 mg/dL), they were 87%, 52%, 1.8, and 0.25, respectively. Glycated hemoglobin level had poorer test characteristics than fasting plasma glucose level or the OGCT. No studies compared the OGCT with International Association of the Diabetes and Pregnancy Study Groups (IADPSG) diagnostic criteria.

Limitations: The lack of a gold standard for confirming GDM limits comparisons. Few data exist for screening tests before 24 weeks’ gestation.

Conclusion: The OGCT and fasting plasma glucose level (at a threshold of 4.7 mmol/L [85 mg/dL]) by 24 weeks’ gestation are good at identifying women who do not have GDM. The OGCT is better at identifying women who have GDM. The OGCT has not been validated for the IADPSG diagnostic criteria.

Primary Funding Source: Agency for Healthcare Research and Quality.

For author affiliations, see end of text.
This article was published at www.annals.org on 28 May 2013.
Review: Screening Tests for Gestational Diabetes

**Table 1. Comparison of Diagnostic Thresholds for Gestational Diabetes Mellitus***

<table>
<thead>
<tr>
<th>Approach</th>
<th>Glucose Load, g</th>
<th>Diagnostic Criteria</th>
<th>Glucose Threshold, mmol/L (mg/dL)</th>
<th>Abnormal Values, n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fasting 1-h 2-h 3-h</td>
<td></td>
</tr>
<tr>
<td>2-step</td>
<td>100</td>
<td>NDDG</td>
<td>5.8 (105) 10.5 (190) 9.1 (165) 8.0 (145)</td>
<td>2</td>
</tr>
<tr>
<td>2-step</td>
<td>100</td>
<td>CC (2000-2010)</td>
<td>5.3 (95) 10.0 (180) 8.6 (155) – 2</td>
<td>2</td>
</tr>
<tr>
<td>2-step</td>
<td>75</td>
<td>ADA (2008)</td>
<td>5.3 (95) 10.0 (180) 8.6 (155) – 2</td>
<td>2</td>
</tr>
<tr>
<td>2-step</td>
<td>75</td>
<td>CDA (2008)</td>
<td>5.3 (95) 10.6 (191) 8.9 (160) – 2</td>
<td>2</td>
</tr>
<tr>
<td>1-step</td>
<td>75</td>
<td>IADPSG</td>
<td>5.1 (92) 10.0 (180) 8.5 (153) – 2</td>
<td>1</td>
</tr>
<tr>
<td>1-step</td>
<td>75</td>
<td>WHO</td>
<td>6.1 (110) – 7.8 (140) – 1</td>
<td>1</td>
</tr>
</tbody>
</table>

*Adapted from Table 1 of reference 14.

The primary objective of this systematic review was to update the 2008 USPSTF review.

**Methods**

The key question for this review was developed by the USPSTF to inform guideline review and development. A technical expert panel that included representatives from the USPSTF and the Office of Medical Applications of Research provided content and methodological expertise. We followed an a priori research protocol for this review. The full technical report is available at [http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=1295&pageaction=displayproduct](http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=1295&pageaction=displayproduct).

**Data Sources and Literature Searches**

A research librarian conducted comprehensive searches from 1995 to May 2012. Databases included Ovid MEDLINE (Appendix Table 1, available at www.annals.org), Ovid MEDLINE In-Process & Other Non-Indexed Citations, the Cochrane Central Register of Controlled Trials, the Cochrane Database of Systematic Reviews, the Database of Abstracts of Reviews of Effects, Global Health, EMBASE, Pascal CINAHL Plus with Full Text (EBSCO host), BIOSIS Previews (Web of Knowledge), Science Citation Index Expanded and Conference Proceedings Citation Index–Science (both via Web of Science), PubMed, Latin American and Caribbean Health Science Literature, the National Library of Medicine Gateway, and OCLC ProceedingsFirst and PapersFirst. We searched trial registries, including the World Health Organization (WHO) International Clinical Trials Registry Platform, ClinicalTrials.gov, and Current Controlled Trials. We also hand-searched proceedings from the scientific meetings (2009–2011) of the American Diabetes Association (ADA), International Association of the Diabetes and Pregnancy Study Groups (IADPSG), International Symposium on Diabetes and Pregnancy, and Australasian Diabetes in Pregnancy Society; searched Web sites of relevant professional associations; and reviewed reference lists of relevant reviews and included studies.

**Study Selection**

Two reviewers independently screened titles and abstracts. Full publications of potentially relevant studies were independently assessed by 2 reviewers using a standardized form. We resolved disagreements by consensus or third-party adjudication.

We included studies if they were English-language prospective studies (that is, trials or cohort studies) that included pregnant women (≥24 or <24 weeks’ gestation) with no known history of preexisting diabetes; reported sufficient data to populate a 2 × 2 table in order to calculate sensitivity and specificity; and compared any GDM screening test (such as blood or urine measurements or a questionnaire) with any reference standard (another screening or diagnostic test). Studies were included regardless of setting and duration of follow-up. The decision to restrict studies to those published in English was made in consultation with the panel of technical experts, who believed that most relevant research would be published in English-language reports.

**Data Extraction and Quality Assessment**

Two reviewers independently assessed the methodological quality of studies and resolved discrepancies by consensus. We assessed studies by using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2) checklist (18). One reviewer used a standardized form to extract data; a second reviewer checked the data for accuracy. Reviewers resolved discrepancies by consensus or third-party adjudication. We extracted study and patient characteristics, inclusion and exclusion criteria, and index test and reference standard characteristics.

**Data Synthesis and Analysis**

We constructed 2 × 2 tables and calculated sensitivity, specificity, and positive and negative likelihood ratios (LRs). Sensitivity and specificity are measures of test accuracy. Likelihood ratios are used to estimate the increased or decreased probability of disease (such as GDM) for a patient and can be used to refine clinical judgment. The larger the positive LR, the greater the accuracy of the test and the greater the likelihood of disease after a positive test...
result; the smaller the negative LR, the smaller the likelihood of disease after a negative test result (19). A positive LR greater than 10 indicates a large and often conclusive probability that the condition is present, whereas a negative LR less than 0.10 suggests a large and often conclusive probability that the condition is not present. An LR of 1 means that a positive or negative result is equally probable in a patient with or without the disease.

If there were more than 3 studies and they were clinically homogeneous (that is, they included women at <24 or ≥24 weeks’ gestation and used similar thresholds and diagnostic criteria), we pooled the data by using a hierarchical summary receiver-operating characteristic curve (HSROC) and bivariate analysis of sensitivity and specificity (20). The HSROC simultaneously compares the sensitivity and specificity (accounting for their correlation) for all studies comparing a particular screening test with GDM diagnostic criteria. We used Review Manager, version 5.0 (The Cochrane Collaboration, Copenhagen, Denmark), to perform meta-analyses and the metandi program in Stata, version 11.0 (StataCorp, College Station, Texas), to fit the bivariate and HSROC models and produce the pooled estimates of sensitivity, specificity, and LRs.

The Results section is organized by type of screening test (for example, OGCT) and is further grouped by the diagnostic criteria used to confirm GDM. We examined the effect of screening before and after 24 weeks’ gestation. Sensitivities, specificities, and LRs and their 95% CIs are presented in summary tables that include all screening tests and diagnostic criteria.

Role of the Funding Source
The Agency for Healthcare Research and Quality (AHRQ) and the USPSTF suggested the initial questions but did not participate in the literature search, data analysis, or interpretation of the results. The AHRQ approved copyright assertion for this manuscript.

RESULTS
From 14 398 citations, 51 prospective cohort studies provided data (Appendix Figure 1, available at www.annals.org) (4–6, 8–11, 16, 17, 21–62). The number of women enrolled in each study ranged from 32 to 9270 (median, 709 women). The mean age of participants was 29 years. Most studies (94%) tested for GDM between 24 and 28 weeks’ gestation. One study tested for GDM before 24 weeks’ gestation (35).

Studies assessed several screening tests, including the 50-g OGCT, measurement of fasting plasma glucose or HbA1c level, and risk factor–based screening. The studies confirmed GDM by using criteria developed by Carpenter and Coustan, ADA (endorsed from 2000–2010), the National Diabetes Data Group (NDDG), WHO, and others. The lack of a gold standard to confirm a diagnosis of GDM limited our ability to compare the results of studies that used different diagnostic criteria. Different criteria resulted in different rates of prevalence, regardless of similarities across study settings and patient characteristics.

We had several concerns about the methodological quality of the studies (Figure 1). For patient selection, 47% of studies were assessed as having high or unclear risk of bias. We had concerns about applicability for this domain, primarily because 55% of studies were conducted in developing countries and used WHO criteria to diagnose GDM. For the reference standard (the criteria used to confirm a diagnosis of GDM), 80% of studies were assessed as having high or unclear risk of bias because the result of the screening test was used to determine whether patients had further testing for GDM (lack of blinding) or this was unclear. The domain of flow and timing was assessed as having low risk of bias in 39% of studies. However, 18% were assessed as having high risk of partial verification bias because not all patients received a confirmatory reference standard if the screening test result was below a certain threshold.

Figure 1. QUADAS-2 risk of bias and applicability assessment, by domain.

QUADAS-2 = Quality Assessment of Diagnostic Accuracy Studies 2.
Screening Tests for Gestational Diabetes

Table 2. Diagnostic Characteristics of Screening Tests for Gestational Diabetes Mellitus

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Studies, n</th>
<th>Screening Test</th>
<th>Criteria</th>
<th>Sensitivity (95% CI), %</th>
<th>Specificity (95% CI), %</th>
<th>LR+ (95% CI)</th>
<th>LR− (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥7.8 mmol/L (≥140 mg/dL)</td>
<td>9</td>
<td>50-g OGCT</td>
<td>CC</td>
<td>85 (76–90)</td>
<td>86 (80–90)</td>
<td>5.9 (4.2–8.3)</td>
<td>0.18 (0.11–0.29)</td>
</tr>
<tr>
<td>≥7.8 mmol/L (≥140 mg/dL)</td>
<td>3</td>
<td>50-g OGCT</td>
<td>ADA (2000–2010)</td>
<td>88 (86–97)*</td>
<td>84 (79–87)*</td>
<td>6.0 (5.1–7.0)*</td>
<td>0.16 (0.06–0.45)*</td>
</tr>
<tr>
<td>≥7.8 mmol/L (≥140 mg/dL)</td>
<td>7</td>
<td>50-g OGCT</td>
<td>NDDG</td>
<td>85 (73–92)</td>
<td>83 (78–87)</td>
<td>5.1 (3.9–6.6)</td>
<td>0.18 (0.10–0.34)</td>
</tr>
<tr>
<td>≥7.8 mmol/L (≥140 mg/dL)</td>
<td>1</td>
<td>50-g OGCT</td>
<td>CDA</td>
<td>81 (58–95)</td>
<td>69 (59–79)</td>
<td>2.6 (1.8–3.8)</td>
<td>0.27 (0.11–0.67)</td>
</tr>
<tr>
<td>≥7.8 mmol/L (≥140 mg/dL)</td>
<td>3</td>
<td>50-g OGCT</td>
<td>WHO</td>
<td>70 (43–85)*</td>
<td>89 (73–94)*</td>
<td>6.5 (5.1–8.3)*</td>
<td>0.33 (0.22–0.52)*</td>
</tr>
<tr>
<td>≥7.2 mmol/L (≥130 mg/dL)</td>
<td>6</td>
<td>50-g OGCT</td>
<td>CC</td>
<td>99 (95–100)</td>
<td>77 (68–83)</td>
<td>4.2 (3.0–5.9)</td>
<td>0.02 (0.003–0.08)</td>
</tr>
<tr>
<td>≥7.2 mmol/L (≥130 mg/dL)</td>
<td>3</td>
<td>50-g OGCT</td>
<td>NDDG</td>
<td>88 (67–90)*</td>
<td>66 (47–84)*</td>
<td>2.7 (1.3–8.9)*</td>
<td>0.14 (0.34–0.55)*</td>
</tr>
<tr>
<td>≥12.2 mmol/L (≥220 mg/dL)</td>
<td>1</td>
<td>50-g OGCT</td>
<td>CC</td>
<td>17 (12–24)</td>
<td>100 (99–100)</td>
<td>Undefined</td>
<td>0.83 (0.78–0.89)</td>
</tr>
<tr>
<td>≥4.7 mmol/L (≥85 mg/dL)</td>
<td>4</td>
<td>Fasting plasma glucose</td>
<td>CC</td>
<td>87 (81–91)</td>
<td>52 (50–55)</td>
<td>1.8 (1.6–2.0)</td>
<td>0.25 (0.16–0.38)</td>
</tr>
<tr>
<td>≥5.0 mmol/L (≥90 mg/dL)</td>
<td>4</td>
<td>Fasting plasma glucose</td>
<td>CC</td>
<td>77 (66–85)</td>
<td>76 (75–77)</td>
<td>3.2 (2.9–3.6)</td>
<td>0.30 (0.20–0.46)</td>
</tr>
<tr>
<td>≥5.1 mmol/L (≥92 mg/dL)</td>
<td>3</td>
<td>Fasting plasma glucose</td>
<td>CC</td>
<td>76 (26–80)</td>
<td>92 (90–95)</td>
<td>7.4 (4.0–13.9)*</td>
<td>0.27 (0.13–0.54)*</td>
</tr>
<tr>
<td>≥3.3 mmol/L (≥59 mg/dL)</td>
<td>5</td>
<td>Fasting plasma glucose</td>
<td>CC</td>
<td>54 (32–74)</td>
<td>93 (90–96)</td>
<td>8.2 (5.9–11.5)</td>
<td>0.49 (0.31–0.79)</td>
</tr>
<tr>
<td>5.0%</td>
<td>1</td>
<td>HbA1c</td>
<td>CC</td>
<td>92 (86–96)</td>
<td>28 (23–33)</td>
<td>1.3 (1.2–1.4)</td>
<td>0.28 (0.15–0.50)</td>
</tr>
<tr>
<td>5.3%</td>
<td>1</td>
<td>HbA1c</td>
<td>IADPSG</td>
<td>12 (7–18)</td>
<td>97 (95–98)</td>
<td>3.9 (2.0–7.7)</td>
<td>0.91 (0.86–0.97)</td>
</tr>
<tr>
<td>5.5%</td>
<td>1</td>
<td>HbA1c</td>
<td>ADA (2000–2010)</td>
<td>86 (72–95)</td>
<td>61 (57–65)</td>
<td>2.2 (1.9–2.6)</td>
<td>0.23 (0.11–0.48)</td>
</tr>
<tr>
<td>7.5%</td>
<td>1</td>
<td>HbA1c</td>
<td>ADA (2000–2010)</td>
<td>82 (72–90)</td>
<td>21 (17–26)</td>
<td>1.0 (0.93–1.2)</td>
<td>0.85 (0.52–1.40)</td>
</tr>
</tbody>
</table>

ADA = American Diabetes Association; CC = Carpenter–Coustan; CDA = Canadian Diabetes Association; HbA1c = hemoglobin A1c; IADPSG = International Association of the Diabetes and Pregnancy Study Groups; LR+ = positive likelihood ratio; LR− = negative likelihood ratio; NDDG = National Diabetes Data Group; OGCT = oral glucose challenge test; WHO = World Health Organization.

* Median (range).

OGCT

Nine studies (4, 22, 23, 27, 31, 33, 38, 56, 63) provided data to estimate sensitivity and specificity of an OGCT using a cut point of 7.8 mmol/L (140 mg/dL); GDM was confirmed by a 100-g OGTT using Carpenter–Coustan criteria. The joint estimates of sensitivity and specificity were 85% and 86%, respectively; the positive and negative LRs were 5.9 and 0.18, respectively (Table 2). Six studies (6, 23, 30, 33, 34, 38) reported results for an OGCT using a cut point of 7.2 mmol/L (130 mg/dL) and confirmed GDM by using the Carpenter–Coustan criteria. The joint estimates of sensitivity and specificity were 99% and 77%, respectively, and the positive and negative LRs were 4.2 and 0.02, respectively (Table 2). If we assume a GDM pretest probability of 5%, the positive LR of 5.9 for the 7.8-mmol/L (140-mg/dL) threshold increases the posttest probability to approximately 24%, compared with a posttest probability of 18% for the 7.2-mmol/L (130-mg/dL) threshold. A negative LR of 0.18 for the former threshold reduces the risk for GDM to 1%; at the latter threshold, the negative LR of 0.02 reduces the probability of GDM to 0.1%. Although certainty in ruling out a diagnosis of GDM is gained with the 7.2-mmol/L (130-mg/dL) threshold, the magnitude of the difference is small enough to be clinically irrelevant unless the pretest probability of GDM is high.

Figure 2 shows 2 HSROCs with the 95% confidence ellipse using pairs of sensitivity and specificity of the studies that provided data for the 2 glucose thresholds. All points are clustered in the upper left-hand quadrant, and the 95% confidence ellipse and diagonal null line do not overlap. This indicates that the ability of the screening test to correctly classify patients with GDM is significantly better than random classification. For the less stringent threshold of 7.8 mmol/L (140 mg/dL), the sensitivity was lower but the specificity was higher, suggesting that the test will result in fewer false-positive results but more false-negative results.

One study (36) assessed an OGCT with a cutoff value of 12.2 mmol/L (220 mg/dL), with GDM confirmed using the Carpenter–Coustan criteria. Sensitivity was 17%, specificity was 100%, and the negative LR was 0.83 (Table 2), thus providing certainty that GDM is present when this threshold is met or exceeded on an OGCT.

The joint estimates of sensitivity and specificity were 85% and 83%, respectively, from the 7 studies (8, 25, 28, 31, 32, 58, 63) that assessed an OGCT with a cut point of 7.8 mmol/L (140 mg/dL) and used the NDDG criteria to confirm GDM (Table 2). Table 2 also summarizes the test characteristics and LRs of the OGCT compared with GDM criteria from the NDDG (≥7.2 mmol/L [≥130 mg/dL]) (8, 26, 49), ADA (75-g glucose dose) (2000–2010 criteria) (35, 51, 55), Canadian Diabetes Association (37), and WHO (21, 29, 32).

One study (n = 749) provided data on screening for GDM in the first and second trimesters; GDM was confirmed using the Japan Society of Obstetrics and Gynecology criteria (35). When the OGCT with a threshold of 7.2 mmol/L (130 mg/dL) was used, the sensitivity and specificity for the first trimester were 93% and 77%, respectively, compared with 100% and 85% for the second trimester. These results should be interpreted cautiously because the women diagnosed with GDM in the first trimester had prepregnancy body mass indices that were significantly higher than those in women who did not have GDM.

Other Tests for GDM

Seven studies (4–7, 24, 38, 52) assessed measurement of fasting plasma glucose level to screen for GDM, which...
was confirmed using Carpenter–Coustan criteria. The studies compared different fasting plasma glucose thresholds and showed a pattern of increasing positive LR as the threshold increased (Table 2). Small increments in fasting plasma glucose level result in clinically significant increases in the probability of GDM being present. Four studies (8–11) evaluated different HbA1c thresholds, with GDM confirmed using different diagnostic criteria; we saw no clear pattern over the range of thresholds (Table 2). Eight studies that examined risk factor–based screening used different diagnostic criteria and could not be pooled (3, 22, 41–43, 46, 59, 62). Sensitivity and specificity varied widely across the studies, and no conclusions could be drawn (Appendix Figure 2, available at www.annals.org). In addition, other less common tests, such as measurement of serum fructosamine and adiponectin, were assessed using different diagnostic criteria. Sensitivity and specificity varied across the screening tests (Appendix Table 2, available at www.annals.org).

**DISCUSSION**

This review included 51 cohort studies that assessed the test characteristics of various screening methods for GDM. The studies used different criteria to confirm a diagnosis of GDM. We found that, in general, when the OGCT with a glucose threshold of 7.2 mmol/L (130 mg/dL) was compared with a threshold of 7.8 mmol/L (140 mg/dL), sensitivity improved but specificity was reduced regardless of the glucose dose and cutoff values used for the OGTT. When the harm of missing a diagnosis (false-negative result) is high, as in women with additional risk factors for adverse pregnancy outcomes, screening tests with high sensitivity are preferred at the expense of specificity. However, if the harm of an incorrect diagnosis (false-positive result) is high, screening tests with high specificity are preferred at the expense of sensitivity. The use of a 12.2-mmol/L (220-mg/dL) cutoff for a diagnosis of GDM on an OGCT is supported by 1 study (36). By accepting a low cutoff for ruling out GDM and a high cutoff for diagnosing GDM on a screening test, the time and cost of a 2-step approach for diagnosis are reduced. Treatment benefits have been shown with a 2-step approach (65, 66).

Measurement of fasting plasma glucose level has been suggested as an alternative to the OGCT. It is more reproducible than post–glucose load testing (67), easier to administer to women who cannot tolerate a glucose drink, and less time-consuming for women and laboratories and has been directly related to pregnancy outcomes (15, 16). Our results show that a fasting plasma glucose test with a threshold of 4.7 mmol/L (85 mg/dL) has sensitivity similar to that of an OGCT. However, its positive LR of 1.8 (vs. 5.9 for the OGCT) suggests that it is not as good at predicting an abnormal OGTT result. Using a threshold of 4.7 mmol/L (85 mg/dL) would result in more women requiring an OGTT unless a high threshold for fasting glucose level (above which no further testing is required) were to be accepted. A fasting plasma glucose level of 5.3 mmol/L or greater (≥95 mg/dL) had good specificity and a positive LR of 8.2 and may be a reasonable threshold above which no further diagnostic testing is required. Although patient preference may be an important consideration in the choice of screening test, it is important to note that there is evidence of population differences in the frequency of fasting or post–glucose load elevations in pregnancy (68). In particular, fasting glucose level did not diagnose GDM as frequently in Asian women as in non-Asian women. Further study is required to confirm whether glucose outcome relationships differ across populations.

Glycated hemoglobin level has poorer test characteristics than fasting plasma glucose level or the OGCT. The use of HbA1c level in pregnant women should not be dismissed because a markedly elevated level may be a quick and simple screening test for the presence of overt diabetes. Further study is required to determine the best HbA1c threshold to detect overt diabetes in pregnant women and whether gestational age–specific thresholds would help identify overt diabetes in this population.

Although we found limited evidence for GDM screening at less than 24 weeks’ gestation, there is clinical justification for early screening in women at high risk for overt

---

**Table 2**

<table>
<thead>
<tr>
<th>Screening Test</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>OGCT</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>Fasting glucose</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>HbA1c</td>
<td>0.6</td>
<td>0.9</td>
</tr>
</tbody>
</table>

**Figure 2. HSROC for the 50-g OGCT at thresholds of 7.8 mmol/L (140 mg/dL) and 7.2 mmol/L (130 mg/dL), with Carpenter–Coustan criteria used to confirm GDM.**

The HSROC with the 95% confidence ellipse graphically compares the sensitivity and specificity for all studies comparing a particular screening test with GDM diagnostic criteria. CC = Carpenter–Coustan; GDM = gestational diabetes mellitus; HSROC = hierarchical summary receiver-operating characteristic curve; OGCT = oral glucose challenge test.
Screening Tests for Gestational Diabetes

Our review did not identify compelling evidence for or against risk factor–based screening. Naylor and colleagues (3) used data from the Toronto Trihospital study to develop a risk scoring system for GDM screening using variable glucose thresholds based on age, body mass index, and race. When the system was applied to a validation group, sensitivity (83%) and specificity (84%) were similar to those of universal screening (3). Adverse pregnancy outcomes associated with GDM are not specific to GDM, and much of the risk for such outcomes is attributable to other factors, such as maternal obesity and excessive maternal weight gain. Variable glucose thresholds based on known risk factors for adverse outcomes would provide a sound scientific approach to GDM screening and may help clinicians align the intensity of clinical care according to patient risk.

The IADPSG has proposed the elimination of a screening test in favor of proceeding directly to a diagnostic test for GDM. A 2-step approach to GDM screening has been shown to be more cost-effective than a 1-step approach (13, 37). Our review did not identify any studies that compared the OGCT with IADPSG criteria.

One of the challenges in comparing studies of screening tests for GDM is the plethora of glucose thresholds and the different glucose loads used for the OGTT (Table 1). The studies in this systematic review assessed the performance of screening tests compared with OGTT results rather than pregnancy outcomes. Ideally, the gold standard comparison for GDM screening tests would be a universally agreed-on set of specific pregnancy outcomes. However, such diagnostic criteria for GDM remain elusive. Although data show a continuous positive relationship between glucose levels and various maternal and neonatal outcomes of varying importance, no clear inflection point exists (15).

We had several concerns about the quality and applicability of the included studies. First, there is concern about partial verification bias in 9 (18%) studies. We conducted sensitivity analyses to assess the effect of these studies on the analyses of the OGCT using Carpenter–Coustan criteria and fasting plasma glucose level at the 4.7-mmol/L (85-mg/dL) and 5.3-mmol/L (95-mg/dL) thresholds. Neither the test characteristics nor our conclusions were affected by inclusion of these studies. Second, 80% of studies were assessed as having high or unclear risk of diagnostic review bias, in which interpretation of the reference standard may have been influenced by the knowledge of the results of the index test. A third concern relates to patient selection and the possibility of spectrum bias; 82% of studies were assessed as having high or unclear concerns about applicability. This was primarily because the studies were conducted in developing countries and used the WHO criteria to diagnose GDM.

Recently published systematic reviews in this area are more limited in terms of study designs included (71, 72) or tests examined (73, 74). A systematic review published in 2010 had a scope similar to that of our review and reached similar conclusions (75). The current systematic review represents an up-to-date and comprehensive summary of existing evidence for all potential approaches to screening for GDM and provides specific recommendations for practice and future research.

The OGCT and measurement of fasting plasma glucose level (at a threshold of 4.7 mmol/L [85 mg/dL]) at 24 weeks’ gestation are good at identifying women who do not have GDM. The OGCT, a glucose load test, is better than the fasting plasma glucose test (4.7 mmol/L [85 mg/dL]) at identifying women who have an abnormal response to larger glucose load tests. Because fasting glucose level better predicts fetal overgrowth (16) and such overgrowth can be modified by metabolic management during pregnancy, a practical option may be to offer women their choice of screening with the OGCT or the fasting plasma glucose test. The diagnostic test endorsed by policymakers for GDM will influence which screening test can be used for GDM because there are no existing comparisons of the OGCT and IADPSG diagnostic criteria. Measurement of HbA1c is not a good screening test for GDM, but further study may demonstrate its potential value for identifying overt diabetes in pregnancy.

From University of Calgary, Calgary, and University of Alberta, Edmonton, Alberta, Canada.

Disclaimer: The findings and conclusions in this document are those of the authors, who are responsible for its content, and do not necessarily represent the views of AHRQ. No statement in this report should be construed as an official position of AHRQ or the U.S. Department of Health and Human Services.

Acknowledgment: The authors thank the following individuals for their contributions to this project: Tamara Durec (searching), Andrea Milne (searching and technical support), Teodora Radisc (article retrieval), Jocelyn Shulhan (screening), Annabritt Chisholm (research support and reference management), and Noah Toppings (Table 1).


Potential Conflicts of Interest: Dr. Donovan: Grant (money to institution): AHRQ, International Diabetes Federation, Eli Lilly. Ms. Muise: Grant (money to institution): AHRQ, Mr. Vandermeer: Grant (money to institution): AHRQ. All other authors have no disclosures. Disclosures can also be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M13-0188.

Requests for Single Reprints: Lois Donovan, MD, Richmond Road Diagnostic and Treatment Centre, 1820 Richmond Road SW, Calgary, Alberta T2T 5C7, Canada; e-mail, lois.donovan@albertahealthservices.ca.
22. 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. [PMID: 16612455]
Current Author Addresses: Dr. Donovan: Richmond Road Diagnostic and Treatment Centre, 1820 Richmond Road SW, Calgary, Alberta T2T 5C7, Canada.
Dr. Hartling: University of Alberta, ECHA 4-472, 11405-87 Avenue, Edmonton, Alberta T6G 1C9, Canada.
Ms. Muise: University of Alberta, Department of Pediatrics, 4-482C Edmonton Clinic Health Academy, 11405-87 Avenue, Edmonton, Alberta T6G 1C9, Canada.
Ms. Guthrie: Alberta Health, 18th Floor, Telus Plaza North Tower, 10025 Jasper Avenue, Edmonton, Alberta T5J 1S6, Canada.
Mr. Vandermeer: University of Alberta, Edmonton Clinic Health Academy, Room 4-496B, 11405-87 Avenue, Edmonton, Alberta T6G 1C9, Canada.
Dr. Dryden: University of Alberta, Department of Pediatrics, 4-474 ECHA, 11405-87 Avenue, Edmonton, Alberta T6G 1C9, Canada.

Author Contributions: Conception and design: L. Donovan, L. Hartling, D.M. Dryden.
Drafting of the article: L. Donovan, L. Hartling, M. Muise, A. Guthrie, D.M. Dryden.
Critical revision of the article for important intellectual content: L. Donovan, L. Hartling, M. Muise, A. Guthrie, B. Vandermeer, D.M. Dryden.
Final approval of the article: L. Donovan, L. Hartling, M. Muise, B. Vandermeer, D.M. Dryden.
Statistical expertise: B. Vandermeer.
Administrative, technical, or logistic support: L. Hartling, A. Guthrie, D.M. Dryden.
Collection and assembly of data: L. Hartling, M. Muise, A. Guthrie, D.M. Dryden.
Appendix Table 1. MEDLINE Search Strategy

<table>
<thead>
<tr>
<th>Ovid MEDLINE (1948 to week 4 of September 2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search date: 9 October 2011</td>
</tr>
<tr>
<td>Results: 8234</td>
</tr>
</tbody>
</table>

1. Diabetes, Gestational/
2. Fetal Macrosomia/
3. Pregnancy Complications/
4. GDM.tw.
5. (gestation$ adj2 (diabet$ or DM or glucose intoleran$ or insulin resistan$)).mp.
6. (pregnan$ adj3 (diabet$ or DM or glucose intoleran$ or insulin resistan$)).mp.
7. (maternal adj2 (diabet$ or DM or glyc?emia or hyperglyc?emia)).tw.
9. macrosomia.tw.
10. or/1-9
11. mass screening/
12. prenatal diagnosis/
13. screen$.tw.
14. ((prenatal or early) adj2 diagnosis).tw.
15. Glucose Tolerance Test/
16. Glucose Intolerance/
17. Blood Glucose/
18. Risk Factors/
19. (glucose adj (tolerance or intolerance or challenge)).tw.
20. OGTT.tw.
21. GCT.tw.
22. (fasting adj2 glucose).tw.
23. or/11-22
24. *Sensitivity and Specificity*/
25. *Predictive Value of Tests*/
26. ROC Curve/
27. specific$.tw.
28. sensitiv$.tw.
29. predictive value.tw.
30. accurac$.tw.
31. diagnostic errors/
32. diagnostic error?.tw.
33. false negative reactions/
34. false positive reactions/
35. (false adj (negative or positive)).tw.
36. *reproducibility of results*/
37. reference values/
38. reference standards/
39. or/24-38
40. and/10,23,39
41. intervention?.mp.
42. (treatment or treatment? or therapy or therapies).mp.
43. manage$.mp.
44. monitor$.mp.
45. exp sulfonylurea compounds/
46. Gliclazide/
47. Glyburide/
48. Tolbutamide/
49. sulfonlyurea?.tw.
50. gliclazid$.tw.
51. glimepirid$.tw.
52. glipizid$.tw.
53. glyburid$.tw.
54. tolbutamid$.tw.
55. (anti-diabet$ or anti-diabet$).tw.
56. insulin?.mp.
57. glibenclamide$.mp.
58. acarbose$.mp.
59. exp Diet Therapy/
60. (diet adj2 (therap$ or restrict$ or advice$)).tw.
61. medical nutrition$ therapy.tw.
62. MNT.tw.
63. exp Life Style/
64. (lifestyle$ or life-style$).mp.
65. Blood Glucose Self-Monitoring/
66. (blood glucose adj (self monitor$ or self-monitor$)).tw.
67. ((self monitor$ or self-monitor$) adj blood glucose).tw.
68. SMBG.tw.
69. Counseling/
70. counsel$.tw.
71. Labor, Induced/
72. (induc$ adj2 labo?).tw.
73. exp Cesarean Section/
74. c?esarean.tw.
75. exp Pregnancy Outcome/
76. pregnanc$ outcome?.tw.
77. or/41-76
78. and/10,77
79. or/40,78
80. clinical trial.pt.
81. randomized controlled trial.pt.
82. randomi?ed.ti,ab.
83. placebo.ti,ab.
84. dt.fs.
85. randomly.ti,ab.
86. trial.ti,ab.
87. groups.ti,ab.
88. or/80-87
89. animals/
90. humans/
91. 89 not (89 and 90)
92. 88 not 91
93. cohort studies/
94. follow-up studies/
95. longitudinal studies/
96. prospective studies/
97. retrospective studies/
98. ((cohort? or follow-up or followup or longitud$ or prospectiv$) adj (study or studies or trial??)).tw.
99. or/93-98
100. 99 not 91
101. exp Guideline/
102. Health Planning Guidelines/
103. (clinical adj2 guideline?).tw.
104. CPG?.tw.
105. ((practice or consensus or position) adj2 (guideline? or recommendation? or statement?)).tw.
106. standard?.tw.
107. protocol?.tw.
108. or/101-107
109. meta analysis.mp,pt.
110. review.pt.
111. search:.tw.
112. or/109-111 [Reviews balanced - HIRU]
113. and/79,92 [Clinical trials & RCTs]
114. and/79,100 [Observational studies]
115. and/79,108 [Guidelines]
116. and/79,112 [SRs MA$]
117. or/113-116
118. limit 117 to (english language and yr = "2000 -Current")
119. limit 117 to (english language and yr = "2000 -2005")
120. limit 117 to (english language and yr = "2006 -Current")
121. remove duplicates from 119
122. remove duplicates from 120
123. or/121-122
124. 112 or 113 or 114 or 115
125. 113 or 114 or 115
126. limit 125 to (english language and yr = "2000 -Current")
127. limit 125 to (english language and yr = "2000 -2005")
128. remove duplicates from 127
129. limit 125 to (english language and yr = "2006 -Current")
130. remove duplicates from 129
131. 128 or 130
132. 113 or 114

Appendix Table 1—Continued
Appendix Table 1—Continued

133. limit 132 to (english language and yr="2000 -Current")
134. limit 132 to (english language and yr="2000 -2005")
135. remove duplicates from 134
136. limit 132 to (english language and yr="2006 -Current")
137. remove duplicates from 136
138. 135 or 137

Appendix Figure 1. Summary of evidence search and selection.

Total citations retrieved from electronic literature searches
($n = 14398$)

References selected for further examination of titles and abstracts
($n = 598$)

Potentially relevant references identified by hand-searching
($n = 30$)
Not retrieved: 8

Full-text articles retrieved and evaluated for inclusion
($n = 620$)

Included
($n = 151$)

Duplicate publications
($n = 26$)

Excluded
($n = 469$)
Ineligible comparator: 227
Duplicate: 10
Intervention: 12
Retrospective cohort (KQ 1): 54
Outcome: 34
Population: 15
Publication type: 106
Study design: 11

Unique studies
($n = 125$)

Excluded during extraction (no comparison or outcome of interest found)
($n = 28$)

Extracted studies
($n = 97$)

Studies addressing objectives of this review*
($n = 51$)

KQ = key question.
* This systematic review was part of a larger technical report. The search was done to identify relevant studies for all objectives of the full report, which is available at http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=H11005&pageaction=displayproduct.
### Appendix Figure 2. Forest plot of sensitivity and specificity of risk factor screening for gestational diabetes, by diagnostic criteria (Carpenter–Coustan, American Diabetes Association [2000–2010], National Diabetes Data Group, and World Health Organization).

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayach et al, 2006 (22)</td>
<td>11</td>
<td>173</td>
<td>2</td>
<td>155</td>
<td>0.85 (0.55–0.98)</td>
<td>0.47 (0.42–0.53)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hill et al, 2005 (42)</td>
<td>42</td>
<td>368</td>
<td>7</td>
<td>368</td>
<td>0.86 (0.73–0.94)</td>
<td>0.50 (0.46–0.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jensen et al, 2003 (43)</td>
<td>100</td>
<td>1798</td>
<td>24</td>
<td>3313</td>
<td>0.81 (0.73–0.87)</td>
<td>0.65 (0.63–0.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ostlund and Hanson, 2003 (46)</td>
<td>29</td>
<td>544</td>
<td>32</td>
<td>3011</td>
<td>0.48 (0.35–0.61)</td>
<td>0.85 (0.83–0.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pöyhönen-Alho et al, 2005 (59)</td>
<td>15</td>
<td>108</td>
<td>4</td>
<td>405</td>
<td>0.79 (0.54–0.94)</td>
<td>0.79 (0.75–0.82)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naylor et al, 1997 (3)</td>
<td>57</td>
<td>240</td>
<td>12</td>
<td>1262</td>
<td>0.83 (0.72–0.91)</td>
<td>0.84 (0.81–0.85)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van Leeuwen et al, 2010 (62)</td>
<td>32</td>
<td>395</td>
<td>11</td>
<td>540</td>
<td>0.74 (0.59–0.86)</td>
<td>0.58 (0.55–0.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wijeyaratne et al, 2006 (41)</td>
<td>134</td>
<td>552</td>
<td>10</td>
<td>157</td>
<td>0.93 (0.88–0.97)</td>
<td>0.22 (0.19–0.25)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Threshold values were author-defined. FN = false-negative; FP = false-positive; TN = true-negative; TP = true-positive.
### Appendix Table 2. Diagnostic Characteristics of Other Screening Tests for Gestational Diabetes Mellitus

<table>
<thead>
<tr>
<th>Test</th>
<th>Study, Year (Reference)</th>
<th>Country</th>
<th>Women, n</th>
<th>Threshold (mmol/L)</th>
<th>Reference Standard</th>
<th>Sensitivity (95% CI), %</th>
<th>Specificity (95% CI), %</th>
<th>LR+ (95% CI)</th>
<th>LR− (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serum fructosamine</td>
<td>Agarwal et al, 2011 (39)</td>
<td>United Arab Emirates</td>
<td>849</td>
<td>≥237 μmol/L</td>
<td>ADA (2000-2010 criteria)†</td>
<td>86 (0.78–0.91)</td>
<td>23 (0.14–0.20)</td>
<td>1.12 (1.03–1.22)</td>
<td>0.61 (0.38–0.97)</td>
</tr>
<tr>
<td>Serum fructosamine</td>
<td>Uncu et al, 1995 (8)</td>
<td>Turkey</td>
<td>42</td>
<td>≥2.85 mmol/L</td>
<td>CC</td>
<td>71 (0.45–0.88)</td>
<td>46 (0.30–0.64)</td>
<td>1.33 (0.83–2.15)</td>
<td>0.62 (0.25–1.54)</td>
</tr>
<tr>
<td>Serum fructosamine</td>
<td>Agarwal et al, 2001 (10)</td>
<td>United Arab Emirates</td>
<td>430</td>
<td>≥210 μmol/L</td>
<td>CC</td>
<td>92 (0.86–0.96)</td>
<td>23 (0.19–0.28)</td>
<td>1.20 (1.11–1.30)</td>
<td>0.34 (0.18–0.65)</td>
</tr>
<tr>
<td>Fasting plasma insulin</td>
<td>Kauffman et al, 2006 (52)</td>
<td>United States</td>
<td>123</td>
<td>≥93 μmol/L</td>
<td>NDDG</td>
<td>56 (0.37–0.73)</td>
<td>71 (0.62–0.79)</td>
<td>1.96 (1.23–3.13)</td>
<td>0.62 (0.39–0.98)</td>
</tr>
<tr>
<td>Fasting plasma insulin</td>
<td>Yachi et al, 2011 (55)</td>
<td>Japan</td>
<td>509</td>
<td>≥36.69 pmol/L</td>
<td>JSOG§</td>
<td>48 (0.43–0.53)</td>
<td>72 (0.63–0.79)</td>
<td>1.71 (1.26–2.34)</td>
<td>0.72 (0.62–0.84)</td>
</tr>
<tr>
<td>Capillary blood glucose</td>
<td>Agarwal et al, 2008 (40)</td>
<td>United Arab Emirates</td>
<td>1662</td>
<td>≥4.9 mmol/L</td>
<td>ADA (FGP)</td>
<td>84 (0.79–0.89)</td>
<td>75 (0.73–0.77)</td>
<td>3.40 (3.05–3.78)</td>
<td>0.21 (0.49–0.29)</td>
</tr>
<tr>
<td>Capillary blood glucose</td>
<td>Balaji et al, 2012 (53)</td>
<td>India</td>
<td>819</td>
<td>≥7.8 mmol/L</td>
<td>WHO</td>
<td>80 (0.71–0.87)</td>
<td>98 (0.97–0.99)</td>
<td>53.5 (29.5–97.0)</td>
<td>0.20 (0.13–0.31)</td>
</tr>
<tr>
<td>Capillary blood glucose</td>
<td>Wijeyaratne et al, 2006 (41)</td>
<td>Sri Lanka</td>
<td>853</td>
<td>≥7.2 mmol/L</td>
<td>WHO</td>
<td>63 (0.54–0.7)</td>
<td>37 (0.34–0.41)</td>
<td>0.99 (0.87–1.15)</td>
<td>1.00 (0.80–1.27)</td>
</tr>
<tr>
<td>Glucose source</td>
<td>Eslamian and Ramezani, 2008 (30)</td>
<td>Iran</td>
<td>138</td>
<td>Breakfast containing 50 g of carbohydrates</td>
<td>ADA (2000-2010 criteria)</td>
<td>83 (0.55–0.95)</td>
<td>86 (0.79–0.91)</td>
<td>5.93 (3.60–9.75)</td>
<td>0.19 (0.06–0.69)</td>
</tr>
<tr>
<td>Glucose source</td>
<td>Lamar et al, 1999 (28)</td>
<td>United States</td>
<td>136</td>
<td>28 jellybeans (50 g)</td>
<td>NDDG</td>
<td>40 (0.12–0.77)</td>
<td>85 (0.78–0.90)</td>
<td>2.66 (0.85–8.38)</td>
<td>0.71 (0.34–1.45)</td>
</tr>
<tr>
<td>Glucose source</td>
<td>Rust et al, 1998 (48)</td>
<td>United States</td>
<td>448</td>
<td>Meal containing 100 g of carbohydrates</td>
<td>ADA (2000-2010 criteria)</td>
<td>25 (0.10–0.50)</td>
<td>98 (0.96–0.99)</td>
<td>12.5 (3.92–39.91)</td>
<td>0.77 (0.58–1.02)</td>
</tr>
</tbody>
</table>

ADA = American Diabetes Association; CC = Carpenter–Coustan; FPG = fasting plasma glucose; IWC = International Workshop-Conference on Gestational Diabetes Mellitus; JSOG = Japan Society of Obstetrics and Gynecology; LR+ = positive likelihood ratio; LR− = negative likelihood ratio; NDDG = National Diabetes Data Group; WHO = World Health Organization.

* To convert to mmol/L, divide by 1000.
† 75-g glucose load.
‡ To convert to pmol/L, multiply by 1 000 000.
§ Fasting plasma insulin level obtained at <13 wk gestation.
|| (Fructosamine level/total protein level) – (glucose level/100).¶ To convert to mg/dL, divide by 0.0555.