Screening for gestational diabetes: a systematic review and economic evaluation

Scott D A, Loveman E, McIntyre L, Waugh N

Authors' objectives
To review research on screening for gestational diabetes mellitus (GDM). The review focused on the diagnostic performance and costs of various screening methods. The authors also assessed GDM screening against the criteria of the UK National Screening Committee.

Searching
MEDLINE, EMBASE, PubMed, the Science Citation Index, the Social Sciences Citation Index, the National Research Register and the Cochrane Library were searched for English language papers from 1966 to 2000; the search terms were reported. The reviewers also examined the citations of retrieved papers and searched their own reference collections.

Study selection
Study designs of evaluations included in the review
Studies of any design were eligible for inclusion.

Specific interventions included in the review
Studies that evaluated any screening test for GDM were eligible for inclusion. The tests investigated in the primary studies included: the glucose challenge test (GCT), with various timings, thresholds of diagnosis and glucose formulations; fasting plasma glucose (FPG), with thresholds ranging from 4.3 to 6.0 mmol/L; random plasma glucose (RPG), with thresholds from 6.0 to 7.0 mmol/L within 2 hours of eating, and 5.6 to 6.4 mmol/L otherwise; urinalysis; glycosylated haemoglobin; fructosamine levels; ultrasound measurement of the foetal abdominal transverse diameter; calculated risk factors; and risk factors combined with GCT screening.

Reference standard test against which the new test was compared
Inclusion criteria for the reference standard were not stated. The reference standards included the 3-hour 100g glucose tolerance test (GTT), the 2-hour 75g GTT, and the 50g GTT. For each of these, a variety of criteria were used to define GDM. In many studies, the reference standard was not given to all women, but only to those with a positive screening test.

Participants included in the review
No inclusion criteria were specified for the participants, but studies assessing the effects of pre-existing diabetes on pregnant women were excluded. The studies included in the review were carried out in a variety of populations, and covered women with a range of socioeconomic backgrounds. Some studies were conducted in patients with a high prevalence of GDM, e.g. in specific ethnic groups. While some studies involved unselected women, others included only those with risk factors for GDM, or those who had already had an abnormal screening test.

Outcomes assessed in the review
Inclusion criteria for the outcomes were not stated. Many of the included studies assessed the sensitivity, specificity and positive predictive value (PPV) of a screening test. Other outcomes included test reproducibility and side effects. In some cases, receiver operating characteristic (ROC) analysis was used to determine the optimum threshold for the test. The sensitivity, specificity and PPV were reported in the review.

How were decisions on the relevance of primary studies made?
Two independent reviewers selected papers for the review on the basis of the titles and abstracts.

Assessment of study quality
The authors did not state that they assessed validity.

Data extraction
Two reviewers extracted the data. The reviewers extracted the sensitivity, specificity and PPV of the screening tests used and, if sufficient data were provided in the primary studies, these calculations were checked. Other data extracted included thresholds for determining GDM, incidence of GDM in the study population, length of gestation, time since eating, and other outcomes as reported in the primary studies. In the summary of results, glucose levels were presented in mmol/L, using a conversion factor if necessary (1mg/dL = 0.0555mmol/L). In addition, equivalent plasma values were given for studies that presented results as whole blood glucose (plasma glucose equals whole blood glucose multiplied by 1.14).

Methods of synthesis
How were the studies combined?
The studies were combined in a narrative summary. In addition, details of each study were tabulated. An assessment of screening for GDM, against the National Screening Committee criteria, was also provided. This was based partially upon the findings of the review.

How were differences between studies investigated?
Within the summary, the studies were grouped according to the screening test under evaluation. Comparisons of screening tests were discussed separately. The authors provided a qualitative discussion of differences between the studies, commenting on factors such as the prevalence of GDM in the population used for a particular study, and thresholds used in the screening tests.

Results of the review
The review included 135 studies (>290,000 patients); not all reported data on diagnostic performance. Most of the studies were case series; some were observational studies, or controlled trials of varying designs.

Urinalysis (glucosuria) (three studies): The sensitivity ranged from 7% (specificity 98%) to 36% (specificity 98%). The specificity ranged from 84% (sensitivity 27%) to 98% (sensitivity 7% and 36%). The PPV ranged from 7 to 27%.

RPG: At a threshold of 7.0mmol/L within 2 hours of eating and 6.4mmol/L otherwise (2 studies), sensitivities of 17% (specificity 99%) and 16% (specificity 96%) were reported. The PPVs were 4.5% and 47%. With a lower threshold of 6.1mmol/L within 2 hours of eating (1 study), the sensitivity was 46% and the PPV 12%. A study using a threshold of 6.0mmol/L within 2 hours of eating found sensitivities of 41 to 58% and specificities of 74 to 96%, depending on the time of day at which the test was performed.

FPG: With a threshold of 5.3mmol/L (1 study), the sensitivity was 48% and the specificity 97.5%. At a threshold of 4.8mmol/L (1 study), the sensitivity was 81% and specificity 76%. At a threshold of 4.9mmol/L (2 studies), sensitivities of 80% (specificity 40%) and 88% (specificity 78%) were reported, with a PPV of 1.3% in one of the studies. With a threshold of 4.3mmol/L (1 study), the sensitivity was 93% and the specificity 38.5%.

GCT: In a study using a threshold of 10.3mmol/L, sensitivity was reported to be 38%, specificity 96% and PPV 79%. At a threshold of 8.3mmol/L (3 studies), the sensitivities were 96% and 74% (specificity 90% for this study) and the PPVs ranged from 24 to 29%. In a study using a threshold of 8.0 mmol/L, the sensitivity was 82%, specificity 88% and PPV 27%, while a study with a threshold of 7.9 mmol/L found a sensitivity of 79%, a specificity of 87% and a PPV of 15%. At a threshold of 7.8 mmol/L (3 studies), the measured sensitivities were 86% (PPV 28%) and 96% (specificity 84% and PPV 25%), while the third study assumed a sensitivity of 100% and found a specificity of 83% and PPV of 15%. One study used ROC analysis to determine the optimal threshold of 7.5mmol/L, which gave a sensitivity of 100%, a specificity of 80% and a PPV of 21%. One study, assessing the reliability of the GCT at the 7.5mmol/L threshold, found that 27% of women would have been missed if only one test was carried out instead of tests on two consecutive days.

Screening with combined GCT and risk factors: Five studies assessed a combination of GCT and risk factors, using a variety of thresholds. At a threshold of 7.2mmol/L and using greater than or equal to 24 or 25 years’ maternal age as a risk factor (3 studies), the sensitivity ranged from 85 to 92%; the PPV in one of these studies was 14%. At a threshold of 7.8mmol/L for women with at least one risk factor (2 studies), the sensitivities were 95% and 86% (with specificity 65% and PPV 23% in one study). In a study using a threshold of 7.9mmol/L, the sensitivity ranged from 53% (specificity 93%) to 88% (specificity 82%) and the specificity ranged from 80% (sensitivity 62%) to 93% (specificity...
53%), depending on the combination of risk factors used. At a threshold of 8.3mmol/L (3 studies), the sensitivity ranged from 62% (using at least 30 years' maternal age alone as the risk factor) to 95% (using an age threshold of at least 24 years); one of the studies reported a specificity 79% (sensitivity 83%), and two reported PPVs of 31% and 32%.

Several studies assessed the performance of alternative sources of glucose; results for the individual studies are reported in the paper. Results from several studies that assessed the value of fructosamine levels were given. Other tests for which results were reported include: fructosamine (four studies), foetal abdominal transverse diameter (one study), and glycosylated haemoglobin. The paper also reports results for the comparison of GCT with other tests in one or two studies, the impact of different diagnostic criteria on test performance, and the psychological impact of screening.

Cost information
The reviewers reported that several studies (nine from the USA and one from Australia) had assessed the cost-effectiveness of GDM screening in terms of the cost per case of GDM detected. These estimates were based on screening with GCT followed by diagnosis with GTT, and varied according to the time of publication, the GCT threshold, the population, and whether GCT was combined with risk factor screening. The estimates varied from US$173 to US$2,733. No UK data were available.

Authors' conclusions
The evidence was insufficient to support the use of routine screening of all women, and the authors suggested instead that a policy of selective screening based on age and weight should be used. The best screening test was considered likely to be the GCT, ideally combined with an FPG. The value of a follow-up diagnostic GTT was considered doubtful. The authors also concluded that GDM screening does not meet all the National Screening Committee criteria.

CRD commentary
This review had broad inclusion criteria for the interventions and participants, and included any study design. No inclusion criteria were specified for the reference standard or outcomes. As a range of studies assessing different aspects of screening was included, the comparability of the studies was somewhat limited. Several electronic databases were searched for studies, and the search terms and dates were reported. Some attempts were made to search for unpublished research, but it is possible that some studies might have been missed. It was unclear whether the authors applied any language restrictions to their search, therefore the potential for language bias cannot be assessed. The study authors were not contacted for further information. Some steps were taken to minimise reviewer errors and bias, in that the selection of papers was performed independently and in duplicate. It was unclear whether the data extraction was also performed independently. The reviewers did not formally assess the validity of the included studies, although they discussed some aspects of study quality in the tables and narrative. For example, they highlighted that some studies used selected populations, and that many only gave the reference standard test to patients who screened positive.

The authors tabulated details of all the included studies. These tables showed that the included studies differed considerably in their focus, as well as in the patient populations included. The use of a narrative summary was therefore appropriate. However, only certain studies were discussed in the text, and it was unclear how these were selected. The reviewers discussed some differences between the studies in the text, e.g. the use of selected or high prevalence populations, or different thresholds. However, the effects of these differences were not assessed in a systematic way. The potential effects of publication and language bias on the results were not considered. As study quality was not assessed, and the authors' conclusions may not have been based on all the included studies, it is difficult to judge the reliability of the conclusions. The authors appropriately highlighted that further research would be required to confirm their conclusions.

Implications of the review for practice and research
Practice: The authors made several suggestions for discussion, pending further research. They stated that when considering screening, the potential harms of a false-positive diagnosis should be taken into account. For this reason, they recommended that a high threshold should be used, to avoid women being wrongly diagnosed. The authors suggested that screening could be a two-stage process in which risk factors (primarily age, weight and ethnicity) are used to screen women, followed by a GCT with a threshold of 8.2 mmol/L. Following a positive GCT, further testing and treatment would only be used if dietary advice failed to reduce glucose levels.
The authors stated that the extent to which glucose levels affect pregnancy outcomes should be determined, and the effects in different ethnic populations should be considered. If there is a continuum of risk, the level at which intervention is required should be determined. This could be achieved by economic evaluation, considering the cost per adverse pregnancy outcome avoided. Different screening tests need to be compared in large, multicentre trials. In particular, trials are required to assess the risks of adverse pregnancy outcomes and the effectiveness of treatment in women who have normal FPG, but abnormal GCTs. In addition, the value of follow-up GTT diagnostic testing should be evaluated. Finally, the cost-effectiveness of screening should be considered, to determine whether screening is necessary and how selective it should be. [A: The authors recommended that the UK HTA Programme should not commission any further research until the results of two large ongoing studies, HAPO and ACOHIS, were available.]

**Funding**

NHS R&D Health Technology Assessment (HTA) Programme, project number 99/09/50.

**Bibliographic details**


**PubMedID**

12433317

**Original Paper URL**

http://www.hta.ac.uk/project.asp?PjtId=1194

**Indexing Status**

Subject indexing assigned by NLM

**MeSH**

Adolescent; Adult; Age Distribution; Cost-Benefit Analysis; Diabetes, Gestational /diagnosis /epidemiology; Female; Great Britain /epidemiology; Humans; Incidence; Mass Screening /economics /methods; Maternal Age; Middle Aged; Pregnancy; Pregnancy Outcome; Prenatal Diagnosis /methods; Risk Assessment; Risk Factors

**AccessionNumber**

12003008030

**Date bibliographic record published**

31/10/2005

**Date abstract record published**

31/10/2005

**Record Status**

This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.