Evidence-based Practice Center Systematic Review Protocol

Project Title: Screening and Diagnosis of Gestational Diabetes Mellitus

I. Background and Objectives for the Systematic Review

Gestational diabetes mellitus (GDM) is defined as any degree of glucose intolerance that either has its onset or first becomes apparent during pregnancy.[1] Disappearance of GDM postpartum is critical, as previously undiagnosed type 2 diabetes can be mistaken for GDM.[2] Current prevalence estimates for GDM range from approximately 1 to 14 percent of pregnancies in the United States, depending on population characteristics, such as ethnicity and clinical status.[3,4] GDM incidence has increased over the past decades, alongside the increase in rates of obesity and type 2 diabetes, and these trends are expected to continue to rise.[5]

GDM is an important public health concern, as impaired glucose tolerance may affect maternal and fetal health outcomes.[5] Mothers may face an increased risk of labor and birth complications, psychological issues, and an increased likelihood of developing diabetes later in life.[6] Risks for the fetus include macrosomia (excessive birth weight) and birth injuries, such as shoulder dystocia, nerve palsies, and fractures. In addition, risk of glucose intolerance and obesity in childhood is associated with GDM.[6]

Much debate exists regarding whether screening for GDM should be offered routinely for pregnancy and, if so, which method of screening should be used. Due to uncertainty regarding whether screening and treatment are beneficial and cost-effective, there is worldwide variance on screening practices.[3] Universal screening of all pregnant women is the most common current practice in the U.S. A cross-sectional study by Gabbe et al. found that 96 percent of obstetricians routinely screen for GDM.[7] In contrast, current guidelines of the American Diabetes Association (ADA) and American College of Obstetrics and Gynecology (ACOG) state the women at low risk for GDM are unlikely to benefit from screening.[1,8] Low risk is defined as women who meet all of the following criteria: age less than 25 years, not from an ethnic group with increased risk for developing diabetes, body mass index of 25 or less, no previous history of glucose intolerance or adverse outcomes associated with GDM, and no known family history of diabetes.[8]

Numerous screening and diagnostic tests have been used to detect high levels of plasma or serum glucose at various stages of pregnancy. Non-challenge blood glucose tests measuring fasting or random plasma glucose or A1c levels may be performed at the first antenatal visit, particularly for women classified as high risk due to personal history of GDM or marked obesity.[1,9] During 24 to 28 weeks' gestation, the most common method of screening for GDM in North America involves a two-step approach, in which patients with abnormal results on a screening test receive a subsequent diagnostic test.[1,8] Typically, a 50 g oral glucose challenge screening test is initially administered. Patients who meet or exceed a screening threshold receive a more involved diagnostic test, the oral glucose tolerance test (OGTT), where a 75 gram (g) or 100 g oral glucose load is administered in a fasting state, and plasma glucose levels are evaluated after 1, 2, or 3 hours. Alternatively, the OGTT diagnostic test may be used directly for high-risk patients, referred to as the one-step method. In many European countries, obstetricians

1

commonly perform selective screening based on risk factors using a blood glucose test, in contrast to the two-step universal screening approach preferred by obstetricians in the U.S.[3]

In addition to the lack of consensus on an optimal screening and diagnostic approach, various thresholds for positive tests have been recommended by the ADA, ACOG, and World Health Organization (WHO).[1,6,8,10] The International Association of Diabetes and Pregnancy Study Groups (IADPSG) recently issued consensus guidelines recommending an OGTT threshold of 5.1 mmol/l at fasting, 10.0 mmol/l at 1 hour post load, and 8.5 mmol/l at 2 hours post load, based on findings from the Hyperglycermia and Adverse Pregnancy Outcome (HAPO) study. They also recommended that GDM be diagnosed with only one glucose value above their recommended thresholds. A recent study found that the IADPSG criteria increased GDM prevalence nearly threefold compared with the current ADA criteria in a high-risk population, which would increase antenatal visits, laboratory workup, and treatment.[11]

Tests for GDM may vary in terms of their performance characteristics, burden on the patient, and cost. The sensitivity and specificity of these tests reflect how correctly they are able to rule in and rule out GDM, thus affecting the clinical utility of the test. Tests with poor performance may result in patient stress, anxiety, and unnecessary investigation and treatment for false positives[12,13] and delayed diagnosis for false negatives. In addition, some tests involve greater burden on patients due to fasting and repeated glucose measurement at various intervals.

Treatment aims to minimize the risk of adverse maternal and child outcomes associated with glucose intolerance in women diagnosed with GDM. First-line treatment for GDM involves diet modification, glucose monitoring, and moderate exercise. When dietary management fails to achieve glucose control, insulin or oral antidiabetic medications may be used.[14] Because clinical uncertainty exists regarding whether treatment to reduce maternal glucose levels decreases the risks associated with GDM, professional associations disagree on screening recommendations.[15] However, many practitioners have decided to adopt some type of screening based on the results of the Australian Carbohydrate Intolerance Study in Pregnant Women, which found decreased risk for severe perinatal outcomes in patients receiving treatment compared to patients receiving routine prenatal care.[3,15] Potential harms of treatment, including small for gestational age (SGA) neonates and maternal stress, must also be considered in weighing the benefits and risks of treatment.

Based on systematic reviews published in 2003 and 2008, the U.S. Preventive Services Task Force (USPSTF) concluded that there was insufficient evidence to recommend for or against routine screening of all pregnant women.[16,17] However, several key studies have been published since the 2008 USPSTF evidence report.[5,18,19] Due to continued uncertainty as to the optimal approach for GDM screening and diagnosis, the Office of Medical Applications of Research (OMAR) and USPSTF have co-sponsored this systematic review to inform consensus meetings and guideline development.

The primary aims of this review are to: a) identify the test properties of screening and diagnostic tests for GDM; b) evaluate the potential benefits and harms of screening at ≥24 weeks and <24 weeks gestation; and c) determine the effects of treatment in modifying outcomes for women diagnosed with GDM. The benefits and harms of treatments will be considered in this review in order to determine the downstream effects of screening on health outcomes. This review will also address several limitations of the previous USPSTF reviews by examining intermediate outcomes for mother, fetus, and child, and considering other criteria for GDM in addition to the standard criteria of the ADA, ACOG, and WHO.

2

I. The Key Questions

The key questions (KQs) to be investigated in this review were developed and finalized by the OMAR and the USPSTF and are presented below.

KQ1: What are the sensitivities, specificities, reliabilities, and yields of current screening tests for GDM? A) After 24 weeks' gestation? B) During the first trimester and up to 24 weeks' gestation?
- Population: Pregnant women (≥24 weeks' gestation and <24 weeks' gestation) without known preexisting diabetes mellitus (DM)
- Interventions: Any screening or diagnostic test, including one-step, two-step, or other approach
- Comparators: Any reference standard
- Outcomes: Sensitivity, specificity, positive predictive value, negative predictive value, reliability (i.e., accuracy), and yield (i.e., proportion of patients who are screened in)
- Timing: Any duration of followup
- Settings: All settings

KQ2: What is the direct evidence on the benefits and harms of screening women (before and after 24 weeks' gestation) for GDM to reduce maternal, fetal, and infant morbidity and mortality?
- Population: Pregnant women (≥24 weeks' gestation and <24 weeks' gestation) without known preexisting DM
- Interventions: Any screening or diagnostic test, including one-step, two-step, or other approach; if diagnosed with GDM, any treatment
- Comparators: No test for GDM
- Outcomes: Maternal, fetal, and infant morbidity and mortality
- Timing: Any duration of followup
- Settings: All settings

KQ3: In the absence of treatment, how do health outcomes of mothers who meet various criteria for GDM and their offspring compare to those who do not meet the various criteria?
- Population: Pregnant women (≥24 weeks' gestation and <24 weeks' gestation) without known preexisting DM who meet different test thresholds for GDM
- Interventions: None
- Comparators: Pregnant women (≥24 weeks' gestation and <24 weeks' gestation) without known preexisting DM who do *not* meet specific test thresholds for GDM
- Outcomes:
  - Maternal
    - Short-term: preeclampsia/maternal hypertension, cesarean delivery (elective and medically indicated), depression, birth trauma, mortality, weight gain
    - Long-term: type 2 DM risk, obesity, hypertension
  - Fetal/neonatal/child

3

- Short-term: macrosomia, shoulder dystocia, clavicular fracture, brachial plexus injury (permanent and transient), birth injury, hypoglycemia, hyperbilirubinemia, mortality
- Long-term: obesity, type 2 DM, transgenerational GDM
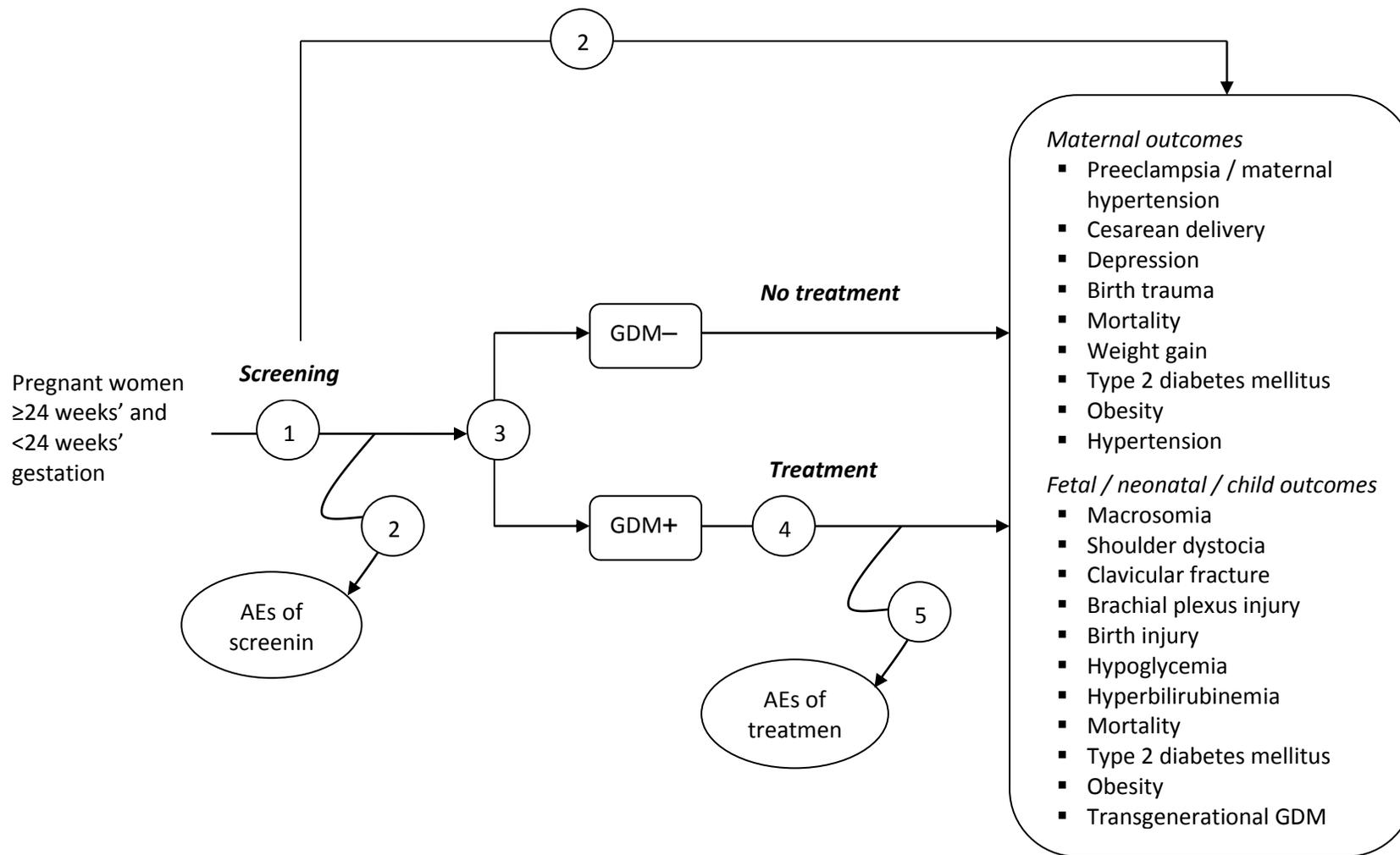- Timing: Any duration of followup
- Settings: All settings

KQ4: Does treatment modify the health outcomes of mothers who meet various criteria for GDM and offspring?
- Population: Pregnant women (≥24 weeks' gestation and <24 weeks' gestation) without known preexisting DM who meet any diagnostic threshold for GDM
- Interventions: Any treatment for GDM including, but not limited to, dietary advice, blood glucose monitoring, insulin therapy, and oral hypoglycemic agents
- Comparators: Placebo or no treatment
- Outcomes:
  - Maternal
    - Short-term: preeclampsia/maternal hypertension, cesarean delivery (elective and medically indicated), depression, birth trauma, mortality, weight gain
    - Long-term: type 2 DM risk, obesity, hypertension
  - Fetal/neonatal/child
    - Short-term: macrosomia, shoulder dystocia, clavicular fracture, brachial plexus injury (permanent and transient), birth injury, hypoglycemia, hyperbilirubinemia, mortality
    - Long-term: obesity, type 2 DM, transgenerational GDM
- Timing: Any duration of followup
- Settings: All settings

KQ5: What are the harms of treating GDM and do they vary by diagnostic approach?
- Population: Pregnant women (≥24 weeks' gestation and <24 weeks' gestation) without known preexisting DM who meet any diagnostic threshold for GDM
- Interventions: Any treatment for GDM including, but not limited to, dietary advice, blood glucose monitoring, insulin therapy, and oral hypoglycemic agents
- Comparators: Placebo or no treatment
- Outcomes: Harms, including anxiety, healthcare system issues, burden on practitioner's office, increased interventions due to treatment bias (e.g., increased cesarean sections resulting from bias of caregivers toward expectation of adverse outcomes), postpartum depression, SGA, costs, and resource allocations
- Timing: Any duration of followup
- Settings: All settings

III. Analytic Framework for Diagnosing Gestational Diabetes Mellitus

Pregnant women ≥24 weeks' and <24 weeks' gestation

Screening

(1)

(3)

(2)

AEs of screenin

No treatment

GDM−

GDM+

Treatment

(4)

(5)

AEs of treatmen

Maternal outcomes
- Preeclampsia / maternal hypertension
- Cesarean delivery
- Depression
- Birth trauma
- Mortality
- Weight gain
- Type 2 diabetes mellitus
- Obesity
- Hypertension

Fetal / neonatal / child outcomes
- Macrosomia
- Shoulder dystocia
- Clavicular fracture
- Brachial plexus injury
- Birth injury
- Hypoglycemia
- Hyperbilirubinemia
- Mortality
- Type 2 diabetes mellitus
- Obesity
- Transgenerational GDM

IV. Methods

This section briefly describes the methods we will use in conducting this review. The methods will be based primarily on the USPSTF Procedure Manual[20] and the Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Comparative Effectiveness Reviews of Medical Interventions.[21]

A. Criteria for Inclusion/Exclusion of Studies in the Review

Publication characteristics

We will include reports of primary research published in English from 1995 to present. This date was selected to incorporate several key studies that were published in the late 1990s. Non-English language studies will be excluded due to lack of translation resources and time restrictions. This decision was made in consultation with the Technical Expert Panel, which expressed no concerns that limiting the search to English language would forfeit important studies. There will be no restriction on study sample size. Conference abstracts will not be included; however, we will contact authors of relevant abstracts in attempt to obtain a published or unpublished manuscript, where available.

Study design

Randomized controlled trials (RCTs), nonrandomized controlled trials (NRCTs), and prospective and retrospective cohort studies with concurrent or nonconcurrent control groups will be eligible for all KQs, including questions examining harms.

Population

We will include studies that examine pregnant women ≥24 weeks' or <24 weeks' gestation who have no known preexisting DM.

Interventions

Studies that used any GDM screening or diagnostic approach, including one-step, two-step, or other approach will be eligible. Because there is no consistent standard for GDM diagnosis, studies will not be excluded based on the diagnostic criteria or threshold used.

For KQs 2, 4, and 5, any treatment for GDM, including but not limited to dietary advice, blood glucose monitoring, insulin therapy (all preparations), and oral hypoglycemic agents will be eligible.

Comparators

Any reference standard will be eligible as a comparator KQ1. Patients who are not screened and patients who meet different criteria for GDM are eligible as comparators for KQ2 and 3, respectively. For KQs 4 and 5, treatment will be compared with placebo or no treatment.

Outcomes

For KQ1, the outcomes of interest are the sensitivity, specificity, positive and negative predictive values, reliability, and yield. For KQ2, outcomes include short- and long-term maternal, fetal, and child morbidity and mortality as well as harms. KQs 3 and 4 will examine a variety of short- and long-term outcomes for the mother and fetus/neonate/child (see above for specific outcomes). KQ5 addresses harms, including anxiety, healthcare system issues, burden on practitioner's office, increased interventions due to treatment bias, postpartum depression,

6

SGA, costs to patients, insurance agencies, and health care systems (see Data Abstraction below), and resource allocation.

B. Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions

In consultation with a medical research librarian, we will systematically search the following electronic resources: MEDLINE® Ovid, Ovid MEDLINE® In-Process & Other Non-Indexed Citations, Cochrane Central Register of Controlled Trials (contains the Cochrane Pregnancy and Childbirth Group, which hand searches journals pertinent to its content area and adds relevant trials to the registry), Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effects (DARE), HealthSTAR, EMBASE, Global Health, CINAHL Plus with Full Text (EBSCO), BIOSIS Previews®, Science Citation Index Expanded® and Conference Proceedings Citation Index- Science (both via ISI Web of Science[SM]), PASCAL, PubMed®, LILACS (Latin American and Caribbean Health Science Literature), National Library of Medicine (NLM) Gateway, OCLC ProceedingsFirst and PapersFirst, and trial registries such as the WHO International Clinical Trials Registry Platform (ICTRP), ClinicalTrials.gov, and Current Controlled Trials. The search will not be peer reviewed.

We will also search the websites of relevant professional associations and research groups including the ADA, IADPSG, International Symposium of Diabetes in Pregnancy, and Diabetes in Pregnancy Society for conference abstracts and proceedings from the past 3 years. We will review the reference lists of relevant reviews (including the 2008 USPSTF review), clinical practice guidelines, and included studies.

The search will be limited to English language RCTs, clinical trials, and cohort studies published from 1995 to date. For the search strategies, a combination of subject headings and keywords will be developed for each electronic resource. Appendix A documents the Medline search strategy. Results from the literature searches will be entered into a Thomson Reuters Reference Manager 11.0.1® bibliographic management database.

We will use a two-step process to select studies. First, two reviewers will independently screen the titles and abstracts (when available) of search results to determine if a study meets the general inclusion criteria. Each report will be rated as "include," "exclude," or "unclear." The full text of all reports classified as "include" or "unclear" will be retrieved for formal review. Next, two reviewers will independently assess the full text of each report using a standard form that outlines the predetermined inclusion and exclusion criteria. We will resolve disagreements through discussion or third-party adjudication, as needed.

C. Data Abstraction and Data Management

We will extract data using a structured, electronic form and enter data into a Microsoft Excel™ 2007 spreadsheet (Microsoft Corp., Redmond, WA). One reviewer will extract data, and a second reviewer will check data for accuracy and completeness. Reviewers will resolve discrepancies by consensus or in consultation with a third party, as needed.

The following data will be extracted from each study: author identification, year of publication, source of funding, study design, population (e.g., inclusion and exclusion criteria, number of patients enrolled, study withdrawals, duration of followup), patient baseline characteristics (e.g., age, race, ethnicity, weight, body mass index, previous diagnosis of GDM, family history of diabetes, comorbidities, smoking prevalence), details of the screening or

diagnostic test and reference standard, glucose threshold for GDM, type of treatment, and outcomes, including adverse events.

We will report outcomes only if quantitative data are reported or can be derived from graphs. We will not include outcomes that are only described qualitatively (e.g., "there was no difference between the groups") or reported only as a p-value in the data analysis.

We will report adverse events as they are reported by the authors of the study. Examples of adverse events that we have identified a priori are listed in KQ5 above. For each adverse event, we will record the number of patients in each group and the number of patients with an adverse event. We will count each event as if it corresponded to a unique individual. Because an individual patient may experience more than one event during the course of the study, this assumption may overestimate the number of patients that experience an adverse event. We will extract any cost-related data, including costs to patients, insurance, or health care system, that is reported in the included studies. However, we will not search for cost effectiveness studies or conduct cost-effectiveness analyses of different treatment strategies. Studies that report only costs and provide no other outcome data will not be included in the review.

When more than one publication reports the results of a single study, we will consider the earliest published report of the main outcome data to be the primary publication. We will extract data from the primary publication first and then any additional outcome data reported in the secondary publications.

D. Assessment of Methodological Quality of Individual Studies

Two reviewers will independently assess the methodological quality of included studies. We will pilot test each tool on a sample of studies and develop decision rules regarding application of the tools a priori through discussions with content and methodology experts. We will resolve discrepancies in quality assessment through consensus or third-party adjudication.

We will assess the methodological quality of studies relevant to KQ1 using the QUADAS checklist.[22] The tool consists of 14 items and assesses important common biases in diagnostic studies including spectrum, incorporation, verification, disease progression, and information biases. Individual items are rated "yes," "no," or "unclear."

For the remaining questions, the internal validity of RCTs and NRCTs will be assessed using the Cochrane Collaboration Risk of Bias tool.[23] This tool consists of seven domains (sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessors, incomplete outcome data, selective reporting, and "other" sources of bias) and a categorization of the overall risk of bias. Each separate domain is rated "low," "unclear," or "high." Blinding and incomplete outcome data will be assessed separately for subjective outcomes (e.g., depression scales) and objective clinical outcomes (e.g., weight gain).

The overall assessment is based on the responses to individual domains. If one or more individual domains are assessed as having a high risk of bias, the overall score will be rated as high risk of bias. The overall risk of bias will be considered low only if all components are rated as having a low risk of bias. The risk of bias for all other studies will be rated as unclear. In addition, information will be collected for each study on the source of funding.[24]

We will assess cohort studies for KQ2–5 using the cohort Newcastle-Ottawa Scale (NOS).[25] The NOS includes eight items assessing sample selection, comparability of cohorts, and the assessment of outcomes. One star is allotted for each item that is adequately addressed in the study, with the exception of the comparability of cohorts, for which a maximum of two stars can

be given. The overall score is calculated by tallying the stars, with a total possible score of nine stars. In addition, information regarding the source of funding will be collected.[24]

For all studies, we will summarize the quality in terms of a "good," "fair," or "poor" overall scores based on the scores of the above tools.

E. Data Synthesis

We will make the following assumptions and perform the following imputations to transform reported data into the form required for analysis. We will extract data from graphs using the measurement tool of Adobe Acrobat 9 Pro (Adobe Systems Inc., California, U.S.) when data are not reported in text or tables. If necessary, we will approximate means by medians and use 95% confidence intervals to calculate approximate standard deviations. We will calculate p-values when they are not reported.[26]

For KQ1, we will construct 2x2 tables and calculate sensitivity, specificity, positive and negative predictive values, reliability and yield of the screening or diagnostic tests. If studies are clinically homogenous, we will simultaneously pool sensitivities and specificities using a hierarchical summary receiver-operator curve.

We will describe the results of studies qualitatively and in evidence tables. We will perform meta-analysis to synthesize the available data when studies are sufficiently similar in terms of their study design, population, screening or diagnostic test, and outcomes. This will be done using the Mantel-Haenszel method for relative risks and the inverse variance method for pooling mean differences. All meta-analyses will use a random effects model. In cases where the outcomes are measured in different scales, we will use a standardized mean difference to pool results, rather than a weighted mean difference.

We will measure statistical heterogeneity among studies using the I^2 statistic. A priori, we will consider an I^2 value of 75 percent or greater to represent substantial heterogeneity, thereby precluding the pooling of studies. In cases of substantial heterogeneity, we will perform subgroup and meta-regression analyses if the number of studies is sufficient to warrant these analyses.[27] Factors we will consider for subgroup analyses include glucose thresholds for tests, type of treatment, maternal age, race or ethnicity, and weight or body mass index, previous diagnosis of GDM, family history of diabetes, and comorbidities, which will be extracted from each study.

We will use Review Manager Version 5.0 (The Cochrane Collaboration, Copenhagen, Denmark) to perform meta-analyses. For dichotomous outcomes, we will compute relative risks to estimate between-group differences. If no event is reported in one treatment arm, a correction factor of 0.5 will be added to each cell of the 2x2 table in order to obtain estimates of the relative risk. For continuous variables, we will calculate mean differences for individual studies. We will report all results with 95% confidence intervals.

Where possible, we will also analyze publication bias both visually using the funnel plot and quantitatively using Begg's[28] and Egger's tests.[29] Review Manager version 5.0.22 (The Cochrane Collaboration, Copenhagen, Denmark) and Stata version 7.0 (Stata Corp., College Station, TX) will be used for all these analyses. In the event that studies cannot be pooled, evidence tables will be produced and a narrative summary of the results will be presented.

F. Grading the Evidence for Each Key Question

Two independent reviewers will grade the strength of the evidence for major outcomes and comparisons using the Evidence-based Practice Center (EPC) GRADE approach[30] and resolve

9

discrepancies by consensus. We will grade the following key outcomes: birth injury, preeclampsia, neonatal hypoglycemia, maternal weight gain, and long-term metabolic outcomes of the child and mother. We will not grade the evidence on the test properties of the screening and diagnostic tests (KQ1). For each outcome, we will assess four major domains: risk of bias (rated as low, moderate, or high), consistency (rated as consistent, inconsistent, or unknown), directness (rated as direct or indirect), and precision (rated as precise or imprecise).

Based on the individual domains, we will assign the following overall evidence grades for each outcome for each comparison of interest: high, moderate, or low confidence that the evidence reflects the true effect. When no studies are available for an outcome or the evidence does not permit estimation of an effect, we will rate the strength of evidence as insufficient.

To determine the overall strength of evidence score, we will first consider the risk of bias domain. RCTs with a low risk of bias will initially be considered to have a "high" strength of evidence, whereas RCTs with high risk of bias and well-conducted cohort studies will receive an initial grade of "moderate" strength of evidence. Low-quality cohort studies will receive an initial grade of "low" strength of evidence. The strength of evidence will then be upgraded or downgraded depending on the assessments of that body of evidence on the consistency, directness, precision, and other optional domains, as applicable.

G. Assessing Applicability

We will assess the applicability of the body of evidence following the PICOTS (population, intervention, comparator, outcomes, timing of outcome measurement, and setting) format used to assess study characteristics. Factors that may potentially weaken the applicability of studies may include study population factors (e.g., race or ethnicity, age, risk level of GDM [i.e., weight, body mass index, previous GDM diagnosis, family history of diabetes], comorbidities), study design (i.e., highly controlled studies [e.g., RCTs] vs. observational studies), setting (e.g., primary vs. tertiary care), and experience of care providers.

V. References

1. Diagnosis and classification of diabetes mellitus. Diabetes Care 2006;29(Suppl 1):S43–8.

2. Hollander M, Paarlberg K, Huisjes AJM. Gestational diabetes: A review of the current literature and guidelines. Obstet Gynecol Surv 2007;62(2):125–36.

3. Berger H, Sermer M. Counterpoint: Selective screening for gestational diabetes mellitus. Diabetes Care 2009;32(7):1352–4.

4. Jovanovic L, Pettitt D. Gestational diabetes mellitus. JAMA 2001;286(20):2516–8.

5. Gillman MW, Oakey H, Baghurst PA, et al. Effect of treatment of gestational diabetes mellitus on obesity in the next generation. Diabetes Care 2010;33:964–8.

6. Dodd J, Crowther C, Antoniou G, et al. Screening for gestational diabetes: The effect of varying blood glucose definitions in the prediction of adverse maternal and infant health outcomes. Aust NZ J Obstet Gynecol 2007;47:307–12.

7. Gabbe S, Gregory R, Power M, et al. Management of diabetes mellitus by obstetrician-gynecologists. Obstet Gynecol 2004;103(6):1229–34.

8. American College of Obstetricians and Gynecologists Committee on Practice Bulletins. ACOG Practice Bulletin. Clinical management guidelines for obstetricain-gynecologists. Number 30, September 2001. Gestational Diabetes. Obstet Gynecol 2011;98(3):525–38.

9. Metzger B, Gabbe S, Persson B, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010;33(3):676–82.

10. Alberti K, Zimmet P. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabet Med 1998;15(7):539–53.

11. Agarwal M, Dhatt G, Shah S. Gestational diabetes mellitus: simplifying the International Association of Diabetes and Pregnancy diagnostic algorithm using fasting plasma glucose. Diabetes Care 2010;33:2018–20.

12. Feig D, Chen E, Naylor C. Self-perceived health status of women three to five years after the diagnosis of gestational diabetes: a survey of cases and matched controls. Am J Obstet Gynecol 1998;178:386–93.

13. Toronto Trihospital Gestational Diabetes Investigators. Cesarean delivery in relation to birth weight and gestational glucose tolerance: pathophysiology or practice style? JAMA 1996;275:1165–70.

14. American Diabetes Association. Position Statement: Gestational diabetes mellitus. Diabetes Care 2003;26:S103–5.

15. Crowther CA, Hiller JE, Moss JR, et al. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med 2005;352(24):2477-86.

16. Hillier T, Vesco K, Pedula K, et al. Screening for gestational diabetes mellitus: A systematic review for the U.S. Preventive Services Task Force. Ann Int Med 2008;148:766–75.

17. U.S. Preventive Services Task Force. Screening for gestational diabetes mellitus: recommendations and rationale. Obstet Gynecol 2003;101(2):393–5.

18. Hyperglycemia and adverse pregnancy outcome (HAPO) study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008;358(19):1991–2002.

19. Landon M, Spong C, Thom E, et al. A multicenter, randomized trial of treatment for mild gestational diabetes. N Eng J Med 2009;361(14):1339–48.

20. U.S. Preventive Services Task Force Procedure Manual. Section 4: Evidence Report Development. AHRQ Publication No. 08-05118–EF. Rockville, Maryland: Agency for Healthcare Research and Quality. July 2008.

21. AHRQ EHCP Methods Guide for Comparative Effectiveness Reviews of Medical Interventions. Rockville, Maryland: Agency for Healthcare Research and Quality. Nov 2010.

22. Whiting P, Rutjes A, Reitsma J, et al. The development of the QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Medical Research Methodology 2003;3:25–38.

23. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available at: http://www.mrc-bsu.cam.ac.uk/cochrane/handbook/front_page.htm. Accessed September 14, 2011.

24. Sismondo S. Pharmaceutical company funding and its consequences: a qualitative systematic review. Contemp Clin Trials 2008;29:109–13.

25. Wells G, Shea B, O'Connell N, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. Department of Epidemiology and Community Medicine, University of Ottawa; 2009.

26. Rutter C, Gastonis C. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001;20:2865–84.

27. Borenstein M, Hedges L, Higgins J, et al. Introduction to meta-analysis. West Sussex: John Wiley & Sons; 2009.

28. Begg C, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994;50:1088–101.

29. Egger M, Smith G, Schneider M, et al. Bias in meta-analysis detected by a single graphical test. Br Med J 1997;315:629–34.

30. Owens D, Lohr K, Atkins D, et al. AHRQ series paper 5: Grading the strength of a body of evidence when comparing medical interventions. Agency for Healthcare Research and Quality and the Effective Healthcare Program. J Clin Epi 2010;63(5):513–23.

VI. Definition of Terms

ACOG	American College of Obstetrics and Gynecology
ADA	American Diabetes Association
DM	Diabetes mellitus
EPC	Evidence-based Practice Center
GDM	Gestational diabetes mellitus
IADPSG	International Association of Diabetes and Pregnancy Study Groups

12

KQ	Key question
NOS	Newcastle-Ottawa Scale
NRCT	Nonrandomized controlled trial
OGTT	Oral glucose tolerance test
OMAR	Office of Medical Applications of Research
RCT	Randomized controlled trial
SGA	Small for gestational age
WHO	World Health Organization
USPSTF	U.S. Preventive Services Task Force

VII. Summary of Protocol Amendments

In the event of protocol amendments, the date of each amendment will be accompanied by a description of the change and the rationale.

VIII. Review of Key Questions

For all EPC reviews, key questions are reviewed and refined as needed by the EPC with input from Key Informants and the Technical Expert Panel (TEP) to assure that the questions are specific and explicit about what information is being reviewed. In addition, for Comparative Effectiveness reviews, the key questions are posted for public comment and finalized by the EPC after review of the comments.

IX. Key Informants

Key Informants are the end users of research, including patients and caregivers, practicing clinicians, relevant professional and consumer organizations, purchasers of health care, and others with experience in making health care decisions. Within the EPC program, the Key Informant role is to provide input into identifying the Key Questions for research that will inform healthcare decisions. The EPC solicits input from Key Informants when developing questions for systematic review or when identifying high priority research gaps and needed new research. Key Informants are not involved in analyzing the evidence or writing the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism.

Key Informants must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals are invited to serve as Key Informants and those who present with potential conflicts may be retained. The Task Order Officer (TOO) and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

X. Technical Experts

Technical Experts comprise a multi-disciplinary group of clinical, content, and methodologic experts who provide input in defining populations, interventions, comparisons, or outcomes as well as identifying particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design and/or methodological approaches

13

do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and recommend approaches to specific issues as requested by the EPC. Technical Experts do not do analysis of any kind nor contribute to the writing of the report and do not review the report, except as given the opportunity to do so through the public review mechanism.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals are invited to serve as Technical Experts and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

XI. Peer Reviewers

Peer Reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodologic expertise. Peer review comments on the preliminary draft of the report are considered by the EPC in preparation of the final draft of the report. Peer reviewers do not participate in writing or editing of the final report or other products. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for Comparative Effectiveness Reviews (CERs) and Technical briefs, be published three months after the publication of the Evidence report.

Potential reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers may not have any financial conflict of interest greater than $10,000. Peer reviewers who disclose potential business or professional conflicts of interest may submit comments on draft reports through the public comment mechanism.

Appendix A: MEDLINE Search Strategy

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.
8. (hyperglyc?emia adj2 pregnan$).tw.
9. macrosomia.tw.
10. or/1-9
11. mass screening/
12. prenatal diagnosis/
13. screen$.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$.tw.
31. diagnostic errors/
32. diagnostic error?.tw.
33. false negative reactions/
34. false positive reactions/
35. (false adj (negative or positive)).tw.
36. "reproducibility of results"/
37. reference values/
38. reference standards/
39. or/24-38
40. and/10,23,39
41. intervention?.mp.
42. (treating or treatment? or therapy or therapies).mp.
43. manage$.mp.
44. monitor$.mp.
45. exp sulfonylurea compounds/
46. Gliclazide/
47. Glyburide/
48. Tolbutamide/
49. sulfonylurea?.tw.
50. gliclazid$.tw.
51. glimepirid$.tw.
52. glipizid$.tw.
53. glyburid$.tw.
54. tolbutamid$.tw.
55. (antidiabet$ or anti-diabet$).tw.
56. insulin?.mp.
57. glibenclamid$.mp.
58. acarbos$.mp.
59. exp Diet Therapy/
60. (diet adj2 (therap$ or restrict$ or advice)).tw.
61. medical nutrition$ therapy.tw.
62. MNT.tw.
63. exp Life Style/

15

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?r).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$ or retrospectiv$) adj (study or studies or trial?)).tw.
99. or/93-98
100. 99 not 91
101. and/79,92 [Clinical trials & RCTs]
102. and/79,100 [Observational studies]
103. or/101-102
104. (comment or editorial or historical article or letter or news or newspaper article).pt.
105. 103 not 104
106. limit 105 to (english language and yr="1995 -2005")
107. limit 105 to (english language and yr="2006 -Current")
108. remove duplicates from 106

16

109. remove duplicates from 107
110. or/108-109