Effective Health Care Program

Future Research Needs Paper
Number 7

Future Research Needs for the Management of Gestational Diabetes

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

This report is based on research conducted by the Johns Hopkins University Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10061-I). The findings and conclusions in this document are those of the author(s), who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care researchers and funders of research make well-informed decisions in designing and funding research and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of scientific judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical research and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances.

This report may be used, in whole or in part, as the basis for research design or funding opportunity announcements. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This information does not represent and should not be construed to represent a determination or policy of the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.

This project was funded under Contract No. 290-2007-10061-I from the Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services (HHS).

Future Research Needs Document

Number 7

Future Research Needs for the Management of Gestational Diabetes

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2007-10061-I

Prepared by:
The Johns Hopkins University Evidence-based Practice Center, Baltimore, MD

Investigators
Wendy L. Bennett, M.D., M.P.H.
Wanda K. Nicholson, M.D., M.P.H., M.B.A.
Ian J. Saldanha, M.B.B.S., M.P.H.
Lisa M. Wilson, Sc.M.
Naomi A. Mckoy, M.S.
Karen A. Robinson, Ph.D.

AHRQ Publication No. 11-EHC005-EF
November 2010

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation: Bennett WL, Nicholson WK, Saldanha IJ, Wilson LM, Mckoy NA, Robinson KA. Future Research Needs for the Management of Gestational Diabetes. Future Research Needs Paper No. 7. (Prepared by Johns Hopkins University under Contract No. 290-2007-10061-I.) AHRQ Publication No. 11-EHC005-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2010. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ) conducts the Effective Health Care Program as part of its mission to organize knowledge and make it available to inform decisions about health care. As part of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Congress directed AHRQ to conduct and support research on the comparative outcomes, clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services to meet the needs of Medicare, Medicaid, and the Children's Health Insurance Program (CHIP).

AHRQ has an established network of Evidence-based Practice Centers (EPCs) that produce Evidence Reports/Technology Assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care. The EPCs now lend their expertise to the Effective Health Care Program by conducting comparative effectiveness reviews (CERs) of medications, devices, and other relevant interventions, including strategies for how these items and services can best be organized, managed, and delivered.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews are useful because they define the strengths and limits of the evidence, clarifying whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about systematic reviews, see http://effectivehealthcare.ahrq.gov/reference/purpose.cfm.

AHRQ expects that CERs will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. In addition, AHRQ is committed to presenting information in different formats so that consumers who make decisions about their own and their family's health can benefit from the evidence.

As part of a new effort in 2010, AHRQ has supported EPCs to work with various stakeholders, including patients, to further develop and prioritize the future research needed by decisionmakers. The Future Research Needs products are intended to inform and support researchers and those who fund research to ultimately enhance the body of comparative effectiveness evidence so that it is useful for decisionmakers.

Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the Web site (www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an email list to learn about new program products and opportunities for input. Comparative effectiveness reviews will be updated regularly.

Acknowledgments

The Johns Hopkins University Evidence-based Practice Center thanks the other authors of the 2008 evidence report on gestational diabetes, stakeholders from Johns Hopkins University, and stakeholders external to Johns Hopkins University (each listed below).

External Inputs

Other Authors of 2008 Evidence Report on Gestational Diabetes who Contributed to this Project

(Contributions included review of list of gaps abstracted from 2008 evidence report with feedback and suggestions for other gaps.)

Kesha Baptiste-Roberts, Ph.D., M.P.H.
Penn State University
Hershey, PA

Shari Bolen, M.D., M.P.H.
Case Western Reserve University and MetroHealth
Cleveland, OH

Tiffany L. Gary-Webb, Ph.D.
Columbia Mailman School of Public Health
New York, NY

Sherita H. Golden, M.D., M.H.S.
Johns Hopkins University School of Medicine
Baltimore, MD

Catherine T. Witkop, M.D., M.P.H.
United States Air Force Academy
Colorado Springs, CO

Local Stakeholders (from Johns Hopkins University)

(Contributions included online and in-person review and feedback on research questions, with suggestions for other research questions. Potential conflicts of interest were disclosed through the use of Disclosures Statements.)

Michele Beaulieu, M.S.W.
Johns Hopkins Bayview Medical Center
Baltimore, MD

Linda Dunbar, Ph.D.
Johns Hopkins HealthCare
Baltimore, MD

Edith D. Gurewitsch, M.D.
Johns Hopkins University School of Medicine
Baltimore, MD

Roxanne Jamshidi, M.D.
Johns Hopkins University School of Medicine
Baltimore, MD

Christine Mckinney, M.S.
Johns Hopkins Bayview Medical Center
Baltimore, MD

Sonal Singh, M.D., M.P.H.
Johns Hopkins University School of Medicine
Baltimore, MD

External Stakeholders (Not from Johns Hopkins University)

(Contributions included online review and feedback on research questions, with suggestions for other research questions. These stakeholders also participated in the Delphi method of formal consensus development. Potential conflicts of interest were disclosed through the use of Disclosures Statements.)

Donald R. Coustan, M.D.
Warren Alpert Medical School of Brown University and Women and Infants Hospital of Rhode Island
Providence, RI

Erica P. Gunderson, Ph.D.
Kaiser Permanente Division of Research
Oakland, CA

Carol J. Homko, Ph.D.
Temple University School of Medicine
Philadelphia, PA

Catherine Kim, M.D., M.P.H.
University of Michigan
Ann Arbor, MI

Janet C. King, Ph.D.
Children's Hospital Oakland Research Institute
Oakland, CA

Jeanne Mahoney, R.N., B.S.N.
American College of Obstetricians and Gynecologists
Washington, DC

Flint Porter, M.D.
University of Utah School of Medicine
Salt Lake City, UT

Samuel F. Posner, Ph.D.
Centers for Disease Control and Prevention
Atlanta, GA

Caroline Signore, M.D., M.P.H.
Eunice Kennedy Shriver National Institute of Child Health and Human Development
Rockville, MD

AHRQ Contacts

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Center Program
Agency for Healthcare Research and Quality

Christine Chang, M.D., M.P.H.
Task Order Officer
Evidence-based Practice Center Program
Agency for Healthcare Research and Quality

Contents

Tables

Figures

Appendixes

Executive Summary

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its comparative effectiveness reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Background

Gestational diabetes mellitus (GDM), the most common medical complication of pregnancy, is defined as carbohydrate intolerance of variable degree, with an onset or first recognition occurring during pregnancy. Studies estimate that GDM affects about 7 percent of births occurring in the United States. GDM is associated with both maternal and neonatal complications. Women with GDM are at high risk for developing noninsulin dependent (type 2) diabetes mellitus.

In 2008, the Johns Hopkins University Evidence-based Practice Center (JHU EPC) completed an Agency for Healthcare Research and Quality (AHRQ) funded evidence report on glucose management, delivery management, postpartum risk assessment, and diagnostic tests for type 2 diabetes in women with GDM. The report focused on the following four key questions (KQs):

Key Question I. What are the risks and benefits of an oral diabetes agent (e.g., glyburide), as compared to all types of insulin, for GDM?

Key Question II. What is the evidence that elective labor induction, cesarean delivery, or timing of induction is associated with benefits or harm to the mother and neonate?

Key Question III. What risk factors are associated with the development of type 2 diabetes after a pregnancy with GDM?

Key Question IV. What are the performance characteristics of diagnostic tests for type 2 diabetes in women with GDM?

The report authors made the following conclusions: (1) maternal glucose levels do not differ substantially in those treated with insulin vs. insulin analogues or oral agents; (2) average infant birth weight may be lower in mothers treated with insulin than with glyburide; (3) induction at 38 weeks may reduce the macrosomia rate, with no increase in cesarean delivery rates; (4) anthropometric measures, fasting blood glucose (FBG), and 2-hour glucose value are the strongest risk factors associated with development of type 2 diabetes; (5) FBG had high specificity, but variable sensitivity, when compared to the 75-gm oral glucose tolerance test (OGTT) in the diagnosis of type 2 diabetes after delivery.

Overall, the evidence was graded either as low strength or insufficient to address the key questions. Because of the widespread deficiencies in the literature, the research team identified broad research gaps and suggested higher quality clinical studies to address each key question. Therefore, the framework for identifying and describing research gaps identified in this report may be unique and most applicable to future reports with uniformly low or insufficient strength of evidence.

In January 2010, AHRQ requested that the JHU EPC develop and pilot test a process to identify research needs. The objective of the project was to help AHRQ establish a standard process for identifying research needs in its evidence reports and to identify research needs for the management of GDM.

Methods

We completed an eight-step process to systematically identify research needs for the management of GDM:

Step 1—Identification and Abstraction of Research Gaps

We reviewed the "Discussion" section and, where necessary, the "Results" section of the 2008 evidence report to identify research gaps. In addition, we abstracted gaps from the five published manuscripts derived from the report. We developed a conceptual model to illustrate the report's key questions.

Step 2—Feedback from Authors of 2008 Evidence Report

We contacted the authors of the report to request their individual feedback on the abstracted gaps.

Step 3—Translation of Research Gaps into Researchable Questions

After receiving feedback from the authors of the report, we revised the gaps and translated them into researchable questions. Questions were generated using the PICO (population, intervention, comparison, and outcomes) framework. We refer to the original report's questions as "key questions" and newly generated researchable questions as "research questions."

Step 4—Online and In-Person Feedback on Research Questions from Local Stakeholders

We invited a group of stakeholders from our institution to provide feedback to refine the research questions. This local group included two obstetricians/gynecologists, one nutritionist/dietitian, one epidemiologist/methodologist, and two members as proxy for the patient/consumer perspective.

Online feedback. For each research question, we asked whether it was worded clearly, the likely clinical benefit/importance of addressing it (on a 9-point scale, higher score indicating greater clinical benefit/importance), and the likely ability for researchers to conduct a study to address it (on the same scale, higher score indicating higher feasibility). These scores helped the team to refine the questions and were provided to help direct the subsequent discussion and refinement of the questions during the in-person meeting. Questions were not removed at this stage.

In-person meeting. The local stakeholders participated in a 1.5-hour in-person meeting. Our purpose for this meeting was to present a summary of the online comments and results, to present the refined research questions, and to solicit further feedback on these questions.

Step 5—Online Feedback, Consensus Development, and Prioritization of Research Questions by External Stakeholders (Delphi Approach)

We invited a group of stakeholders from various institutions across the United States. This external group included two obstetricians/gynecologists, two nutritionists/dietitians, two epidemiologists/methodologists, two research funders, and two members as proxy for the patient/consumer perspective.

Delphi approach. We used the Delphi method for consensus development, deciding *a priori* that this would be repeated until consensus is reached, with no more than three rounds. In the opening round, stakeholders were asked whether each research question was worded clearly and asked to rate the likely clinical benefit/importance of each. We used these ratings to classify the clinical benefit/importance of each research question as:

- *high* (between 7 and 9),
- *medium* (between 4 and 6), or
- *low* (between 1 and 3).

We defined consensus as at least 75 percent of stakeholders rating clinical benefit/importance within a single category. Research questions with consensus were not retained for following rounds. For each remaining research question in the next two rounds, we provided a summary (mean, range) of the ratings of clinical benefit/importance and a synopsis of comments from the previous round.

Step 6—Prioritization of Outcomes for Key Questions I and II

Key Questions I and II had numerous outcomes. For each of these key questions, we asked external stakeholders to rank the top three outcomes essential to include in a clinical trial to address the research questions.

Step 7—Refinement of Final Research Questions and Development of Conceptual Models To Display Research Gaps

For each key question, we developed a conceptual model to illustrate the research questions of high clinical benefit/importance.

Step 8—Evaluation of Process

We asked contributors to the project (authors of the 2008 evidence report, local stakeholders, and external stakeholders) to evaluate our process and the final research questions.

Results

Step 1—Identification and Abstraction of Research Gaps

Gaps included research in populations of diverse races/ethnicities and GDM types (diet-controlled, oral medications-requiring, insulin-requiring); and interventions and outcomes from the 2008 report. In addition to gaps specific to a key question, we abstracted general research gaps for GDM (e.g., interventions to improve compliance with postpartum screening for type 2 diabetes).

Step 2—Feedback from Authors of 2008 Evidence Report

Five of the eight contacted authors provided the following feedback on the identified gaps. Authors suggested that more randomized controlled trials (RCTs), or for populations in whom RCTs would be particularly challenging, larger well-designed observational studies should be conducted. Authors also stressed the need for consistent definitions of outcomes across studies.

Step 3—Translation of Research Gaps into Researchable Questions

We incorporated the feedback from the authors of the 2008 report and translated the gaps into seventeen research questions (see Appendix B). Of these questions, six, three, five, and three related to Key Questions I, II, III, and IV, respectively.

Step 4—Online and In-Person Feedback on Research Questions from Local Stakeholders

Online feedback. We received online feedback from all local stakeholders. The ratings are summarized below:

- *Key Question I (six research questions).* Research Question I-1 (effectiveness and safety of any second generation sulfonylurea vs. any insulin) and research question I-2 (effectiveness and safety of metformin vs. any insulin) were rated as highest clinical benefit/importance. Research Question I-3 (comparative effectiveness and safety of any oral hypoglycemic medication [i.e., a second generation sulfonylurea or metformin] vs. any insulin) was rated most feasible for a study to be conducted.
- *Key Question II (three research questions).* Research Question II-1 (effectiveness and safety of elective labor induction vs. expectant management at term) was rated as highest clinical benefit/importance. Research Question II-3 (effectiveness and safety of elective labor induction or cesarean delivery vs. expectant management at term in women with insulin-requiring [class A2] GDM) was rated the most feasible for a study to be conducted.
- *Key Question III (five research questions).* Research Question III-1 (maternal health behaviors (e.g., breastfeeding, physical activity) as risk factors for type 2 diabetes) was rated as highest clinical benefit/importance. Research Question III-4 (maternal co-morbidities [e.g., obesity, hypertension, hypercholesterolemia] as risk factors for type 2 diabetes) was rated the most feasible for a study to be conducted.

- *Key question IV (three research questions).* Research Question IV-1 (accuracy of a single FBG test vs. the full 2-hour 75-gm OGTT) was rated as highest clinical benefit/importance and most feasible for a study to be conducted.

In-person meeting. Through discussions 4 questions were removed. In each case, other questions were modified to include the concepts and issues addressed in those questions. Six questions were added at this step. The wording was modified in each of the remaining questions.

Step 5—Online Feedback, Consensus Development, and Prioritization of Research Questions by External Stakeholders (Delphi Approach)

The refined list included nineteen research questions: four, two, nine, and four related to Key Questions I, II, III, and IV, respectively (see Appendix C). We have described below the results from each round, highlighting the sixteen research questions that achieved consensus (low, medium or high clinical benefit/importance) and the three that did not achieve consensus after three Delphi rounds.

Delphi round 1. We included nineteen questions in round 1. Consensus of high clinical benefit/importance was established on the following eight questions:

- *I-1.* Effectiveness and safety of any second generation sulfonylurea vs. any insulin
- *I-2.* Effectiveness and safety of metformin vs. any insulin
- *II-1.* Effectiveness and safety of elective labor induction at 40 weeks vs. expectant management
- *II-2.* Effectiveness and safety of cesarean delivery at 40 weeks vs. expectant management
- *III-1.* Maternal health behaviors (e.g., breastfeeding, physical activity) as risk factors for type 2 diabetes
- *III-7.* Family history, gene mutations, genotypes, gene-environment interactions, epigenetic modifications, or other biomarkers as risk factors for type 2 diabetes
- *III-8.* Comparative effectiveness of various lifestyle interventions (e.g., diet, physical activity) for prevention of type 2 diabetes mellitus, glucose intolerance/impaired fasting glucose, and obesity
- *IV-3.* Comparative effectiveness of strategies or interventions to improve clinician compliance with postpartum screening guidelines for type 2 diabetes.

Delphi round 2. Eleven questions did not achieve consensus in round 1. Of these, six achieved consensus in round 2.

Consensus of high clinical benefit/importance was established on the following five questions:

- *I-3.* Comparative effectiveness and safety of various insulin regimens in terms of type/duration, dosing, and frequency of administration
- *III-3.* Maternal metabolic measures (e.g., fasting insulin levels, OGTT measures) as risk factors for type 2 diabetes
- *III-4.* Co-morbid conditions (e.g., advanced maternal age, obesity, hypertension) as risk factors for type 2 diabetes

- *III-9.* Comparative effectiveness of various educational and behavioral change strategies (e.g., patient education about diabetes risk, lactation support) for prevention of type 2 diabetes
- *IV-2.* Performance characteristics (sensitivity, specificity, and reproducibility) of the HbA1c test vs. the 2-hour 75-gm OGTT in screening for type 2 diabetes.

Consensus of medium clinical benefit/importance was established the following question:

- *III-6.* Interconception interval as a risk factor for type 2 diabetes.

Delphi round 3. Five questions did not achieve consensus in round 2. Of these, the following two questions achieved consensus high clinical benefit/importance in round 3:

- *I-4.* Effectiveness and safety of other hypoglycemic drug classes (e.g., thiazolidinediones) vs. any insulin or other hypoglycemic drugs
- *IV-1.* Performance characteristics (sensitivity, specificity, and reproducibility) of a single FBG test vs. the 2-hour 75-gm OGTT in screening for type 2 diabetes.

Research questions for which no consensus on clinical benefit/importance was established after three Delphi rounds.

- *III-2.* Maternal psychosocial factors (e.g., mood disorders, substance use disorders) as risk factors for type 2 diabetes
- *III-5.* Contraceptive method (e.g., progestin-only) as a risk factor for type 2 diabetes
- *IV-4.* Comparative effectiveness of health information technology interventions to track postpartum screening for type 2 diabetes in women with a history of GDM.

Step 6—Prioritization of Outcomes for Key Questions I and II

Key Question I. Chronic diseases (e.g., obesity, type 2 diabetes) in the offspring was rated the highest among thirty outcomes. The next most important were hypertensive disorders of pregnancy (e.g., gestational hypertension, preeclampsia), medication adherence, and large for gestational age and macrosomia.

Key Question II. Cesarean delivery (including primary cesarean and repeat cesarean) and indication for cesarean delivery was rated the highest among nineteen outcomes. The next most important were birth trauma (e.g., bone fractures, brachial plexus palsy) and neonatal intensive care unit (NICU) admission.

Step 7—Refinement of Final Research Questions and Development of Conceptual Models to Display Research Gaps

Final research questions. Through the three rounds, fifteen research questions achieved consensus of high clinical benefit/importance (see Table A) and one achieved consensus of medium clinical benefit/importance (overall consensus establishment rate=84.2 percent).

Table A. Results from step 7 (final list of research questions rated as high clinical benefit/importance)

Sr. No.	Ques. Number	Final Research Questions	Clinical Benefit/ Importance*** (Mean, Range on a Scale of 1-9)
1	I-1	What are the effectiveness and safety of any of the second generation sulfonylureas compared to any insulin in the treatment of gestational diabetes with regard to the following short- and long-term maternal outcomes, neonatal outcomes, and long-term offspring outcomes?*	Mean=8.2 Range=7 to 9
2	I-2	What are the effectiveness and safety of metformin compared to any insulin in the treatment of gestational diabetes with regard to the following short- and long-term maternal outcomes, neonatal outcomes, and long-term offspring outcomes?*	Mean=7.9 Range=6 to 9
3	I-3	What are the comparative effectiveness and safety of various insulin regimens in terms of type/duration, dosing, and frequency of administration in the treatment of gestational diabetes with regard to the following short- and long-term maternal outcomes, neonatal outcomes, and long-term offspring outcomes?*	Mean=7.3 Range=6 to 9
4	I-4	What are the effectiveness and safety of other hypoglycemic drug classes (e.g., thiazolidinediones, DPP-4 inhibitors, GLP-1 agonists, meglitinides) compared to any insulin or other hypoglycemic drugs in the treatment of gestational diabetes with regard to the following short- and long-term maternal outcomes, neonatal outcomes, and long-term offspring outcomes?*	Mean=6.9 Range=4 to 9
5	II-1	What are the effectiveness and safety of elective labor induction at 40 weeks compared to expectant management in women with gestational diabetes with regard to the following maternal and neonatal outcomes?** Populations of Interest: • All women with gestational diabetes • Women with insulin-requiring (class A2) gestational diabetes • Obese women with gestational diabetes • Women with gestational diabetes with high estimated fetal weight (e.g., >4000 or >4500 gram) • Women with different parities • Women of various races/ethnicities.	Mean=7.8 Range=6 to 9
6	II-2	What are the effectiveness and safety of elective cesarean delivery at 40 weeks compared to expectant management in women with gestational diabetes with regard to the following maternal and neonatal outcomes?** Populations of Interest: • All women with gestational diabetes • Women with insulin-requiring (class A2) gestational diabetes • Obese women with gestational diabetes • Women with gestational diabetes with high estimated fetal weight (e.g., >4000 or >4500 gram) • Women with different parities • Women of various races/ethnicities.	Mean=7.3 Range=4 to 9
7	III-1	What is the evidence that maternal health behaviors (e.g., breastfeeding, physical activity, diet) are associated with the risk of developing type 2 diabetes or glucose intolerance/ impaired fasting glucose following a pregnancy with gestational diabetes?	Mean=8.1 Range=7 to 9
8	III-3	What is the evidence that maternal metabolic measures (e.g., fasting insulin levels, OGTT measures, HPA axis stress (subclinical hypercortisolism)) are associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Mean=7.3 Range=6 to 8

Table A. Results from step 7 (final list of research questions rated as high clinical benefit/importance) (continued)

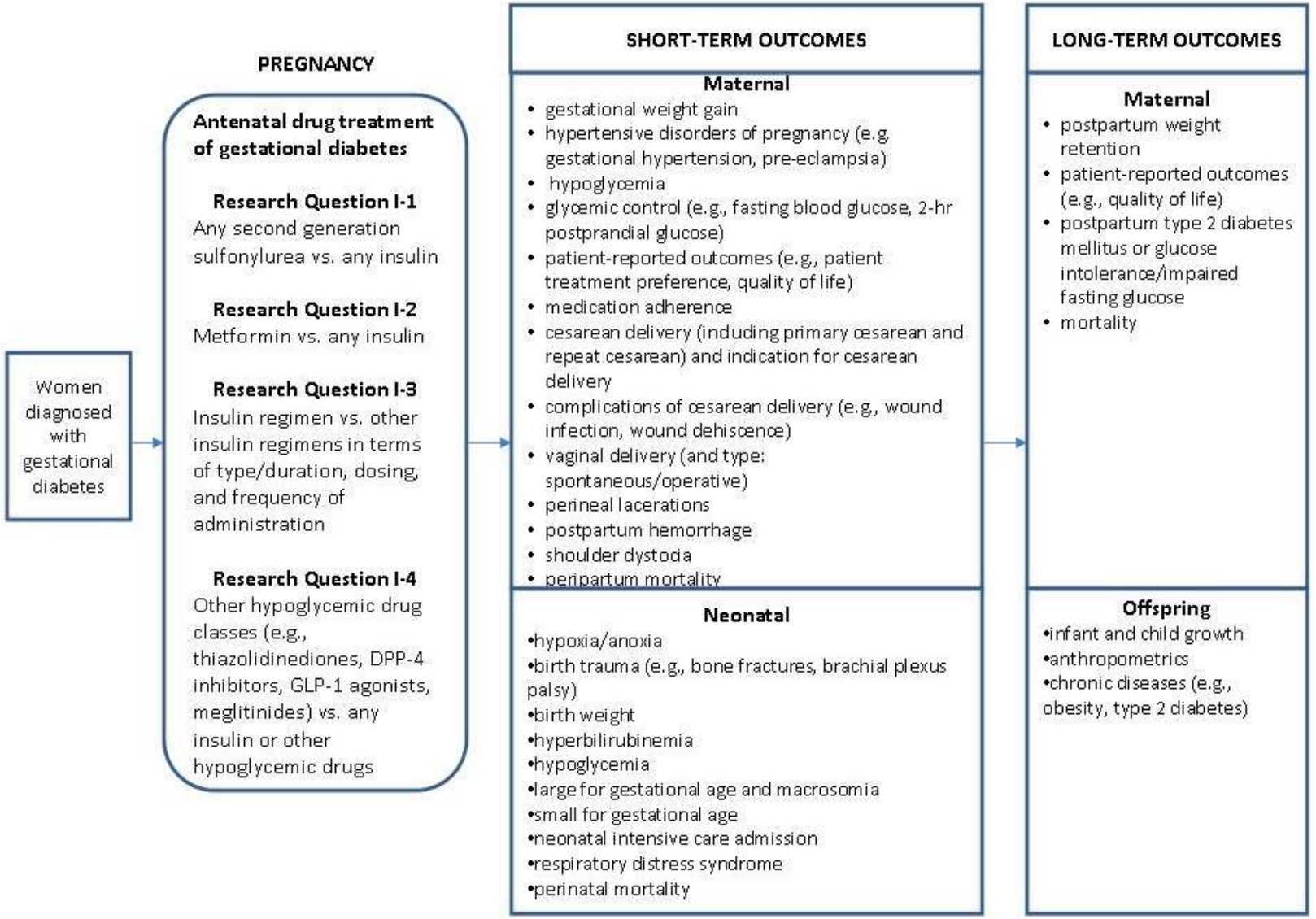
Sr. No.	Ques. Number	Final Research Questions	Clinical Benefit/ Importance*** (Mean, Range on a Scale of 1-9)
9	III-4	What is the evidence that co-morbid conditions (e.g., advanced maternal age, obesity, hypertension, hypercholesterolemia) are associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Mean=7.4 Range=6 to 9
10	III-7	What is the evidence that family history, gene mutations, genotypes, gene-environment interactions, epigenetic modifications, or other biomarkers are associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose among women with gestational diabetes? Are there differences in these associations by race or ethnic group?	Mean=7.4 Range=3 to 9
11	III-8	What is the comparative effectiveness of various lifestyle interventions (e.g., diet, physical activity, smoking) for prevention of type 2 diabetes, glucose intolerance/impaired fasting glucose, and obesity in women with a history of gestational diabetes?	Mean=7.7 Range=6 to 9
12	III-9	What is the comparative effectiveness of various educational and behavioral change strategies (e.g., patient education about diabetes risk, lactation support, diet, physical activity) for prevention of type 2 diabetes and glucose intolerance/impaired fasting glucose in women with a history of gestational diabetes?	Mean=7.3 Range=2 to 9
13	IV-1	What are the performance characteristics (sensitivity, specificity, and reproducibility) of a single fasting blood glucose test compared to the full 2-hour 75-gm OGTT in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes? Does the accuracy of the fasting blood glucose test compared to the full 2-hour 75-gm OGTT vary with the postpartum testing interval in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Mean=6.7 Range=1 to 9
14	IV-2	What are the performance characteristics (sensitivity, specificity, and reproducibility) of the HbA1c test compared to the 2-hour 75-gm OGTT in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes? Does the accuracy of the HbA1c test compared to the full 2-hour 75-gm OGTT vary with the postpartum testing interval in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Mean=7.9 Range=6 to 9
15	IV-3	What is the comparative effectiveness of strategies or interventions to improve clinician compliance with postpartum screening guidelines for type 2 diabetes and glucose intolerance/impaired fasting glucose in women with a history of gestational diabetes?	Mean=7.8 Range=5 to 9

Table A. Results from step 7 (final list of research questions rated as high clinical benefit/importance) (continued)

* **Outcomes for Research Questions I-1, I-2, I-3, and I-4:** **Short-Term Maternal Outcomes:** gestational weight gain, hypertensive disorders of pregnancy (e.g., gestational hypertension, preeclampsia), hypoglycemia, glycemic control (e.g., fasting blood glucose, 2-hr postprandial glucose), patient-reported outcomes (e.g., patient treatment preference, quality of life), medication adherence, cesarean delivery (including primary cesarean and repeat cesarean) and indication for cesarean delivery, complications of cesarean delivery (e.g., wound infection, wound dehiscence), vaginal delivery (and specify type: spontaneous or operative), perineal lacerations, postpartum hemorrhage, shoulder dystocia, and peripartum mortality **Long-Term Maternal Outcomes:** postpartum weight retention, obesity, patient-reported outcomes (e.g., quality of life), development of postpartum type 2 diabetes or glucose intolerance/impaired fasting glucose, and mortality **Neonatal Outcomes:** hypoxia/anoxia, birth trauma (e.g., bone fractures, brachial plexus palsy), birth weight, hyperbilirubinemia, hypoglycemia, large for gestational age and macrosomia, small for gestational age, neonatal intensive care admission, respiratory distress syndrome, and perinatal mortality **Long-Term Offspring Outcomes:** infant and child growth, anthropometrics, and chronic diseases (e.g., obesity, type 2 diabetes).
** **Outcomes for Research Questions II-1 and I-2:** **Maternal Outcomes:** cesarean delivery (including primary cesarean and repeat cesarean) and indication for cesarean delivery, complications of cesarean delivery (e.g., wound infection, wound dehiscence), vaginal delivery (spontaneous, operative), perineal lacerations, hemorrhage, patient-reported outcomes (e.g., patient preference, quality of life), length of hospital stay, pulmonary embolism, and mortality **Neonatal Outcomes:** hypoxia/anoxia, birth trauma (e.g., bone fractures, brachial plexus palsy), birth weight, hyperbilirubinemia, hypoglycemia, large for gestational age and macrosomia, small for gestational age, neonatal intensive care admission, respiratory distress syndrome, and perinatal mortality.
***We used the external stakeholders' ratings of clinical benefit/importance to classify each research question as follows: • *high* clinical benefit/importance (between 7 and 9), • *medium* clinical benefit/importance (between 4 and 6), and • *low* clinical benefit/importance (between 1 and 3). We defined consensus to have been achieved if at least 75 percent (7 out of 9) of stakeholders rated clinical benefit/importance within a single category (high, medium, or low).
Abbreviations: DPP-4=dipeptidyl peptidase-4, GLP-1=glucagon-like peptide-1, HbA1c=hemoglobin A1c, HPA=hypothalamo-pituitary-adrenal, OGTT=oral glucose tolerance test.

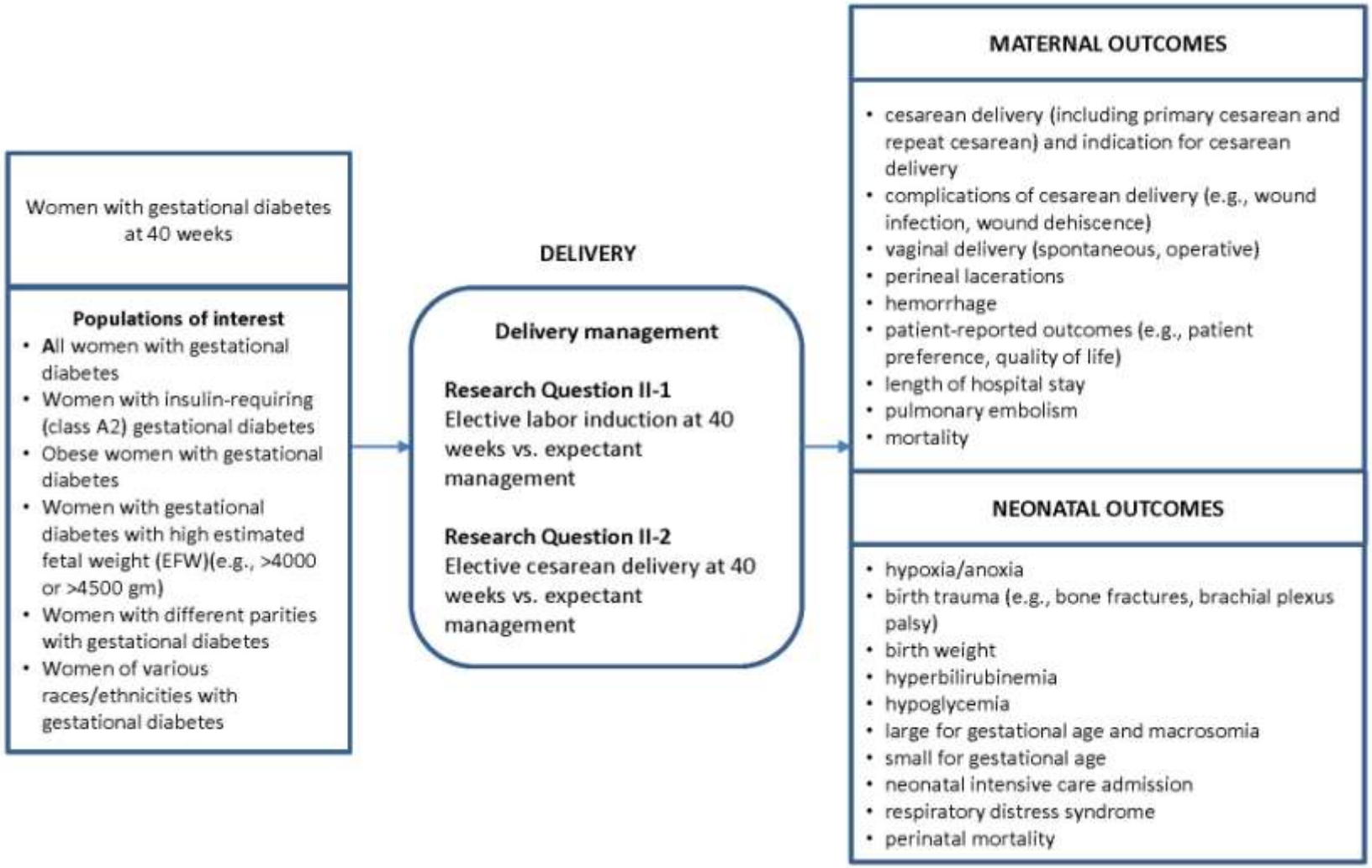
Conceptual Models: See figures A, B, C, and D for conceptual models displaying research questions related to key questions I, II, III, and IV, respectively.

Figure A. Results from step 7 (conceptual model to display research questions related to key question I (KQ-I))



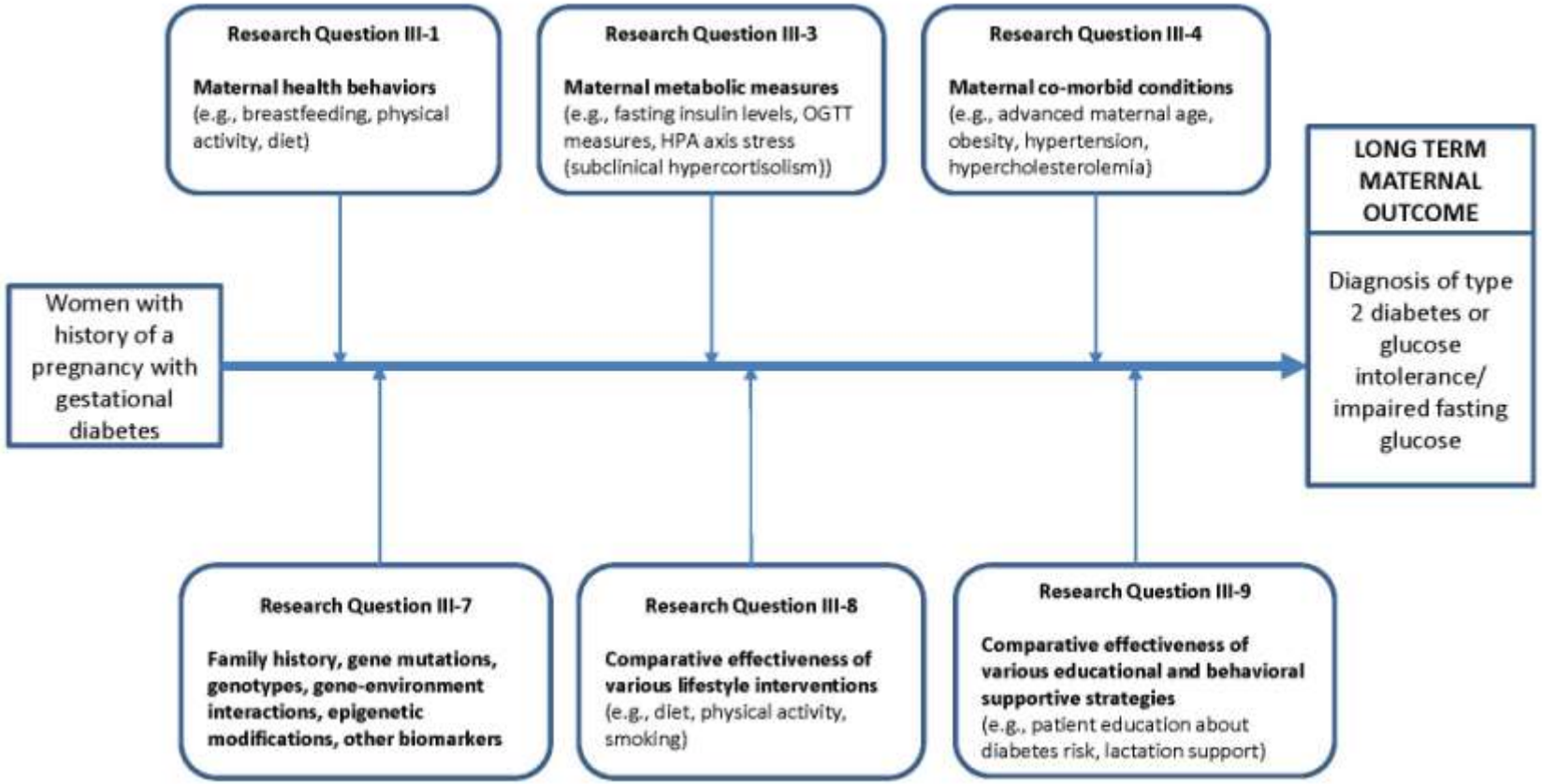
Abbreviations: DPP-4=dipeptidyl peptidase-4, GLP-1=glucagon-like peptide-1.

Figure B. Results from step 7 (conceptual model to display research questions related to key question II (KQ-II))



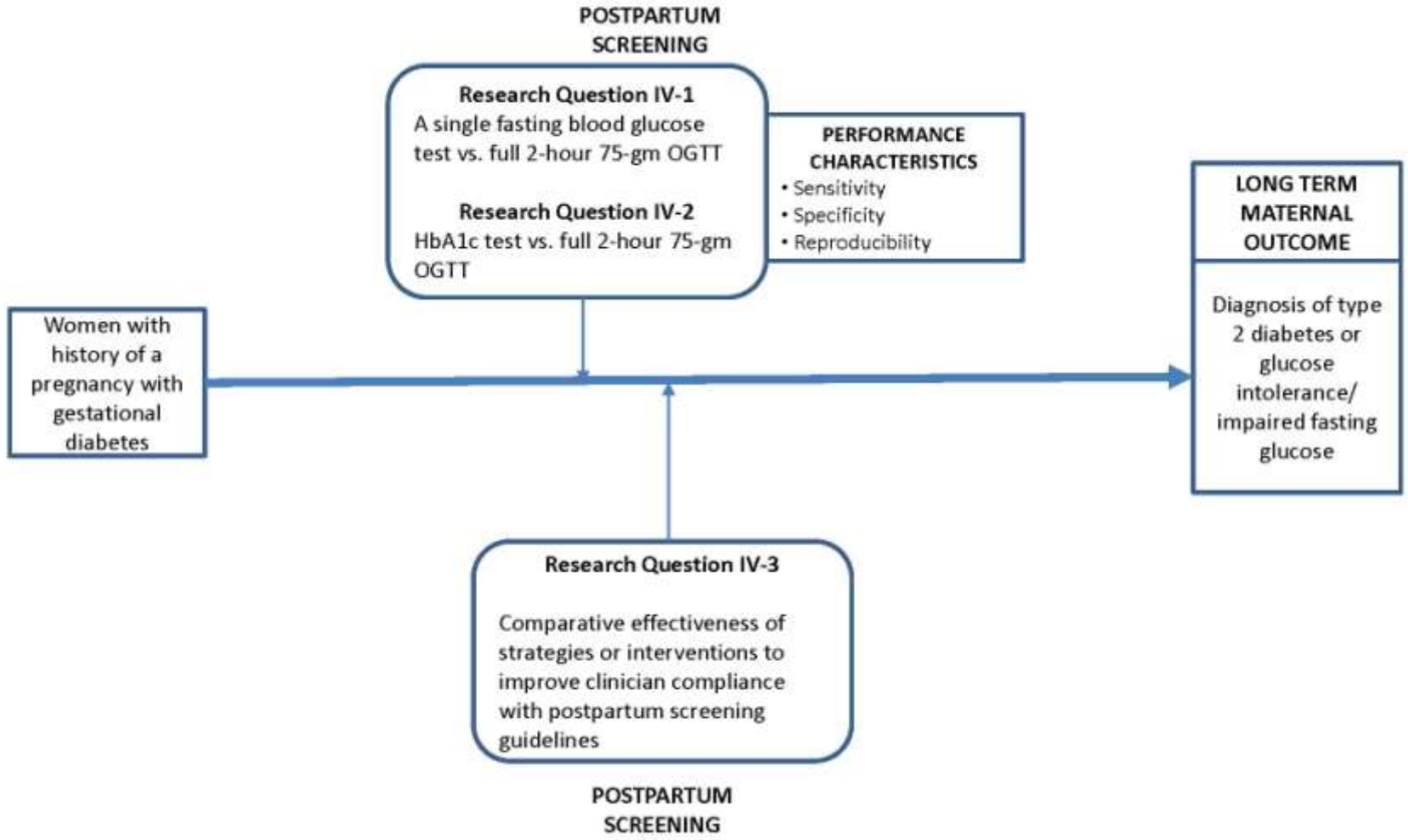
Abbreviations: EFW=estimated fetal weight.

Figure C. Results from step 7 (conceptual model to display research questions related to key question III (KQ-III))



Abbreviations: HPA=hypothalamo-pituitary-adrenal, OGTT=oral glucose tolerance test.

Figure D. Results from step 7 (conceptual model to display research questions related to key question IV (KQ-IV))



Abbreviations: HbA1c=hemoglobin A1c, OGTT=oral glucose tolerance test.

Step 8—Evaluation of Process

All contributors felt that we had accomplished our objective of identifying important research questions for GDM. Additional suggested stakeholders included endocrinologists/diabetologists and patients with current/past GDM. A web-based form was the most preferred mode of participation.

Conclusions

Using the 2008 JHU EPC evidence report on the management and postpartum followup of GDM, we developed an eight-step process for identifying and prioritizing clinically important research needs, with key input from a diverse group of stakeholders. The research needs, reflecting the breadth of the original key questions, address a variety of interventions, risk factors for development of GDM and outcomes. Questions that were specifically added through the involvement with stakeholders included the role of genetics in the development of GDM and the role of lifestyle changes in prevention, as well as questions about appropriate ways to increase patient and physician education and compliance. There did not appear to be a particular type of question that did not reach consensus or was not rated as of high clinical importance. Through this process we propose a final list of fifteen research questions, and high priority outcomes of interest, which highlight the most up-to-date research needs in the field.

There are several strengths to the process we developed. First, our research team had diverse expertise. Three of the members of the team were also part of the original report's research team, one of which was the original report's principal investigator. Second, we invited stakeholders from a wide range of relevant disciplines to ensure a balanced and broad perspective on research needs for GDM. Third, we used the Delphi method to achieve formal consensus development.

There are several limitations to our process for identifying research gaps. In particular, we had limited input from patients and patient liaisons. In addition, this was a fairly resource-intense process, with eight steps, including three Delphi rounds. It is possible that the same research team that just completed an evidence report may not be able to commit the time and resources to these steps. Rather, we would suggest an independent or combination independent-current evidence report team carry out this additional process.

Lessons Learned and Future Directions

Several stakeholders agreed that more interactive approaches, such as Webinars would have been more effective in engaging multiple stakeholders from various geographical regions. We also identified the need to involve a large number of non–research-oriented clinician and patient perspectives in the process, especially related to studying patient-oriented outcomes. For this pilot we recognized certain logistical barriers to involving patients (e.g., institutional review board review, patient selection). Future projects will need to incorporate patients and their perspectives, particularly for prioritizing patient-centered outcomes. Our research team determined that each step was useful, often iterative, and built upon the previous step. However, it may be that certain steps could be abbreviated. The impact of any changes should be evaluated. We anticipate that our process could be used as is, or modified based on the evaluations, as a model for using other evidence reports to take the next step of developing research questions in areas of highest clinical importance.

We identified 15 research needs (see Table A) that were considered by a diverse group of stakeholders to be of high clinical importance. These research needs may be used by professional organizations, researchers and others as a basis for the development of a research agenda.

Background

Gestational diabetes mellitus (GDM), the most common medical complication of pregnancy, is defined as carbohydrate intolerance of variable degree, with an onset or first recognition occurring during pregnancy.[1] Population-based studies estimate that GDM affects about 200,000 (7 percent) of the over 4 million births occurring annually in the United States.[2] GDM is associated with both maternal and neonatal complications.[3-5] Furthermore, women with GDM are at high risk for developing non-insulin dependent (type 2) diabetes mellitus. In a large Canadian cohort, almost 20 percent of women with GDM developed type 2 diabetes within nine years of pregnancy.[6]

In 2008, the Johns Hopkins University Evidence-based Practice Center (JHU EPC) completed an Agency for Healthcare Research and Quality (AHRQ) funded evidence report on glucose management, delivery management, postpartum risk assessment, and diagnostic tests for type 2 diabetes in women with GDM.[7] The report focused on the following four key questions (see Table 1): (I) What are the risks and benefits of an oral diabetes agent (e.g., glyburide), as compared to all types of insulin, for GDM?; (II) What is the evidence that elective labor induction, cesarean delivery, or timing of induction is associated with benefits or harm to the mother and neonate?; (III) What risk factors are associated with the development of type 2 diabetes after a pregnancy with GDM?; (IV) What are the performance characteristics of diagnostic tests for type 2 diabetes in women with GDM?

Table 1. Original key questions from 2008 evidence report on gestational diabetes

Key Question I (KQ-I) What is the evidence for the risks and benefits of oral diabetes agents (e.g., second-generation sulfonylureas and metformin), as compared to all types of insulin, for both the mother and neonate in the treatment of women with gestational diabetes? a. How does maternal outcome vary based on the level of glucose at the initiation of a medication? b. How does neonatal outcome vary based on the level of glucose at the initiation of a medication? Maternal outcomes: cesarean delivery, glycemic control (FBG, 1-hr and 2-hr postprandial glucose (PPG)), hemorrhage, hypoglycemia, operative vaginal delivery, perineal tears, preeclampsia, weight Neonatal outcomes: anoxia, birth trauma, birth weight, congenital malformations, hyperbilirubinemia, hypoglycemia, LGA, macrosomia, mortality, neonatal intensive care admissions, respiratory distress syndrome, shoulder dystocia, SGA.
Key Question II (KQ-II) What is the evidence that elective cesarean delivery or the choice of timing of induction in women with gestational diabetes results in beneficial or harmful maternal and neonatal outcomes? a. What is the evidence for elective cesarean delivery at term, as compared to an attempt at vaginal delivery (spontaneous or induced) at term, with regard to beneficial or harmful maternal and neonatal outcomes in gestational diabetes? i. cesarean vs. spontaneous labor and vaginal delivery ii. cesarean vs. induced labor and vaginal delivery iii. cesarean vs. any attempt at vaginal delivery at term b. What is the evidence for labor induction at 40 weeks, as compared to labor induction at an earlier gestational age (less than 40 weeks) or spontaneous labor, with regard to beneficial or harmful maternal and neonatal outcomes in gestational diabetes? i. labor induction at less than 40 weeks vs. labor induction at 40 weeks ii. labor induction at 40 weeks vs. spontaneous labor iii. labor induction at less than 40 weeks vs. spontaneous labor c. How is the EFW related to outcomes of management of gestational diabetes with elective cesarean delivery or the timing (i.e., gestational age range) of labor induction? d. How is gestational age related to outcomes of management of gestational diabetes with elective cesarean delivery or the choice of timing (i.e., gestational age range) of labor induction? Maternal outcomes: cesarean delivery, hemorrhage, infection, operative vaginal delivery, perineal tears Neonatal outcomes: anoxia, birth trauma, birth weight, congenital malformations, hyperbilirubinemia, hypoglycemia, LGA, macrosomia, mortality, neonatal intensive care admissions, respiratory distress syndrome, shoulder dystocia, SGA.
Key Question III (KQ-III) What risk factors, including but not limited to family history, physical activity, pre-pregnancy weight, and gestational weight gain, are associated with short-term and long-term development of type 2 diabetes following a pregnancy with gestational diabetes?
Key Question IV (KQ-IV) What are the performance characteristics (sensitivity, specificity, and reproducibility) of tests for diagnosing type 2 diabetes after pregnancy in patients with a history of gestational diabetes? Are there differences in the performance characteristics of the test results based on subgroup analysis?

Abbreviations: EFW=estimated fetal weight, FBG=fasting blood glucose, LGA=large for gestational age, PPG=postpartum glucose, SGA=small for gestational age

The report authors identified 45 articles and made the following conclusions: (1) maternal glucose levels do not differ substantially in those treated with insulin vs. insulin analogues or oral agents; (2) average infant birth weight may be lower in mothers treated with insulin than with glyburide; (3) induction at 38 weeks may reduce the macrosomia rate, with no increase in cesarean delivery rates; (4) anthropometric measures, fasting blood glucose (FBG), and 2-hour glucose value are the strongest risk factors associated with development of type 2 diabetes; (5) FBG had high specificity, but variable sensitivity, when compared to the 75-gm oral glucose tolerance test (OGTT) in the diagnosis of type 2 diabetes after delivery.

Overall, the evidence was graded either as low strength or insufficient to address the key questions. Because of the widespread deficiencies in the literature, the authors noted that it was

challenging to identify very specific gaps to target future research. Rather, the research team identified broad research gaps and suggested higher quality clinical studies to address each key question. Therefore, the framework for identifying and describing research gaps identified in this report may be unique and most applicable to future reports with uniformly low or insufficient strength of evidence.

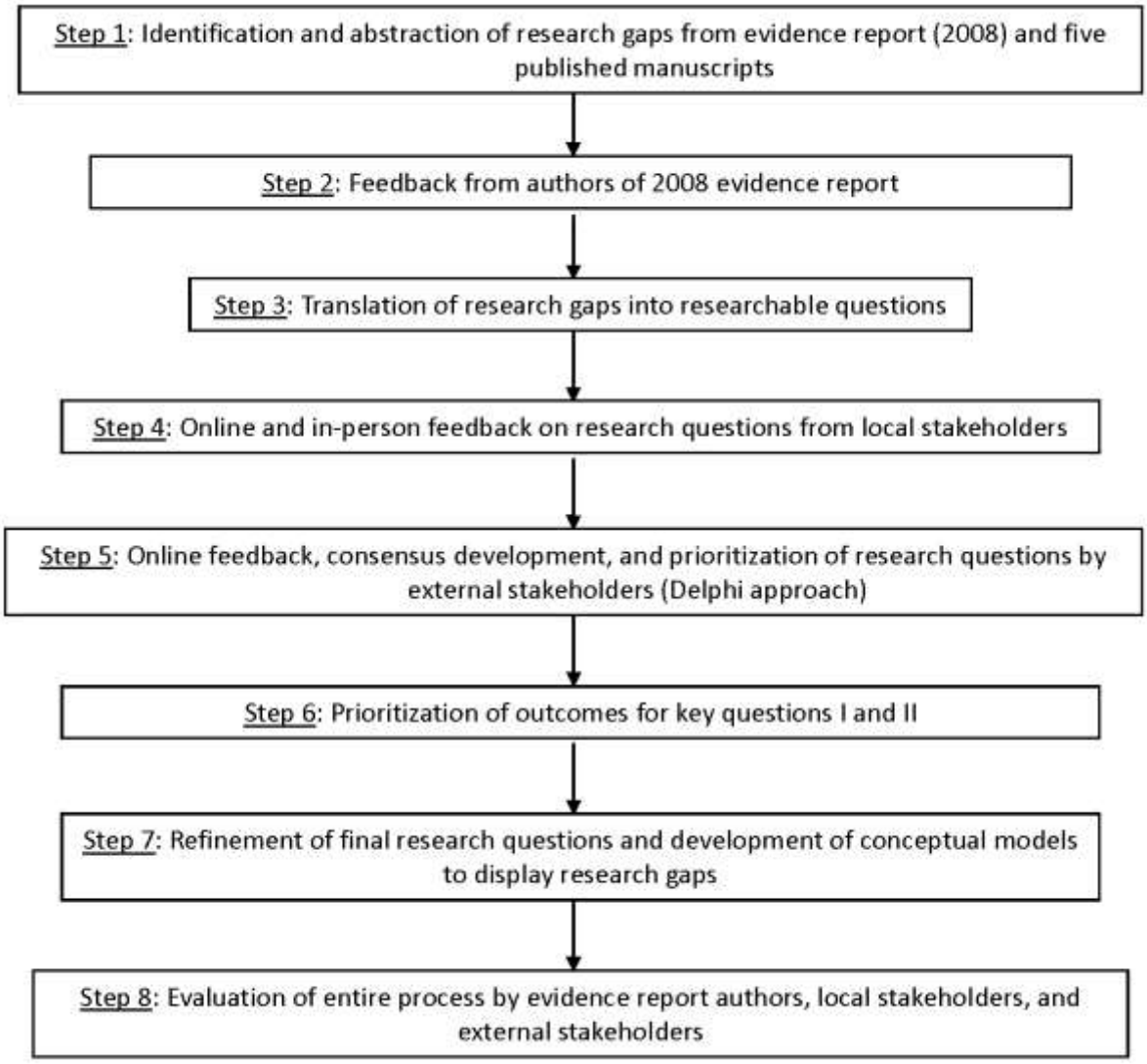
Objective

In January 2010, AHRQ requested that the JHU EPC develop and pilot test a process to identify research needs. The objectives of the project were: (1) to help AHRQ establish a standard process for identifying research needs in its evidence reports and (2) to identify and prioritize future research needs for the management of gestational diabetes.

Methods

The key steps we completed are outlined in Figure 1 and described below:

Figure 1. Steps in project



Step 1—Identification and Abstraction of Research Gaps from Evidence Report (2008) and Five Published Manuscripts

We used the 2008 evidence report to identify gaps in the evidence for the management of GDM.[7] Two authors independently abstracted gaps from the report using custom-designed data abstraction forms in Excel spreadsheets (Microsoft,™ Redmond, WA). Research gaps were abstracted from the "Discussion" section of the report, which includes the "Summary of Key Findings," "Conclusions," "Future Research," and "Implications" sections. Where necessary, authors referred back to the "Results" section of the report. The lists of research gaps abstracted by the two authors were compared and a combined list was created based on consensus. In addition, one author abstracted additional research gaps from the five published manuscripts that

were derived from the report.[8-12] Information about gaps identified in the manuscripts was used to add to or clarify the items in the list of gaps abstracted from the report.

We developed a conceptual model to identify and illustrate the overall location of each of the original report's key questions and the main types of outcomes.

Step 2—Feedback from Authors of 2008 Evidence Report

Three of the eleven authors of the 2008 evidence report on GDM are members of the research team for this project. We contacted the eight other authors to request their individual review and feedback on the list of gaps. For each key question we organized research gaps by population, intervention and comparisons, outcomes, settings, and study designs. We asked the authors to provide their feedback as well as any specific clarifications or additional gaps using Word documents (Microsoft,™ Redmond, WA). We contacted these eight authors via email.

Step 3—Translation of Research Gaps into Researchable Questions

After receiving feedback, clarifications, and additional gaps from the authors of the 2008 evidence report we revised the list of gaps as appropriate. These gaps were then translated into a list of researchable questions. These research questions were developed using the PICO (population, intervention, comparison, and outcomes) framework. The use of this structured framework facilitated the further development and refinement of the specific elements needed for a well-defined research question. The list of research questions was reviewed by all members of the research team to ensure accuracy and completeness.

It is worth clarifying that we refer to the questions from the original report as "key questions" and newly generated researchable questions in this project as "research questions."

Step 4—Online and In-Person Feedback on Research Questions from Local Stakeholders

We invited a local group of stakeholders from our own institution to provide feedback on the research questions we developed (see Table 2). This group comprised six members representing four perspectives with broad clinical, research, and patient advocacy expertise (two physicians [obstetricians and gynecologists], one nutritionist/dietitian, one epidemiologist/methodologist, and two members that served as proxy for the patient/consumer perspective). These stakeholders were contacted via email. We obtained a Disclosure Statement from each stakeholder to ensure that potential conflicts of interest were disclosed. The list of local stakeholders and their Disclosure Statements were approved by AHRQ.

Table 2. Composition of groups of local and external stakeholders

Areas of Expertise	Number of Local Stakeholders (from Johns Hopkins)	Number of External Stakeholders (not from Johns Hopkins)
Physicians (Obstetricians and Gynecologists)	2	2
Nutritionists/Dietitians	1	2
Epidemiologists/Methodologists	1	2
Research Funders	-	2
Patient/Consumer Representatives	2	2
Total	6	9*

* One external stakeholder served as both a nutritionist/dietitian and an epidemiologist/methodologist. Therefore, there were 9 external stakeholders.

Online Feedback

Each local stakeholder was asked to provide feedback on each research question using an online tool (SurveyGizmo,™ Widgix LLC, Boulder, CO). For each research question, we asked the stakeholders to comment on:

- whether or not the research question was worded clearly, and if not, to suggest changes;
- the likely clinical benefit/importance of addressing the research question (on a 9-point Likert scale with higher score indicating greater clinical benefit/importance); and
- the likely ability for researchers to conduct a study to address the research question (on a 9-point Likert scale with higher score indicating higher feasibility).

These scores helped the team to refine the questions and were provided to help direct the subsequent discussion and refinement of the questions during the in-person meeting. The scores helped to illuminate any differences in the rating of clinical importance and were used to help clarify the wording and scope of the questions. Questions were not removed at this stage. In addition, we asked stakeholders to provide overall suggestions or clarifications. We then refined the research questions as appropriate.

In-Person Meeting To Provide Additional Feedback

The same group of local stakeholders was invited to participate in an in-person meeting which lasted 1.5 hours. The purpose of this meeting was to present a summary of the comments and results from the online feedback process, to present the refined list of research questions (organized by key question), and to solicit further feedback on these research questions. In addition to reviewing the research questions from the online feedback, we asked stakeholders to consider study design needs and challenges for each of the research questions to gain an interactive exploration of these issues.

Step 5—Online Feedback, Consensus Development, and Prioritization of Research Questions by External Stakeholders (Delphi Approach)

We invited an external group of stakeholders from various other institutions across the United States (see Table 2). This group comprised representatives of five perspectives (two physicians [obstetricians and gynecologists], two nutritionists/dietitians, two epidemiologists/methodologists, two research funders, and two members that served as proxy for the patient/consumer perspective). These stakeholders were contacted via email. One of the

external stakeholders was deemed to have expertise in two areas (nutrition/dietetics and epidemiology/methodology) and thus represented both. As a consequence, there were nine stakeholders. We believe that inviting stakeholders from a range of relevant disciplines ensured a balanced and broad perspective on research needs for GDM. We obtained a Disclosure Statement from each external stakeholder to ensure that potential conflicts of interest were disclosed. The list of external stakeholders and their Disclosure Statements were approved by AHRQ.

Delphi Approach

We used the Delphi method for consensus development.[13] An online tool (SurveyGizmo,™ Widgix LLC, Boulder, CO) was used to carry out the process. We decided a priori that this iterative process would be repeated until consensus is reached, with the number of rounds not to exceed three. In the opening Delphi round, stakeholders were provided the list of research questions and for each research question were asked to comment on:

- whether or not the research question was worded clearly, and if not, to suggest changes; and
- the likely clinical benefit/importance of addressing the research question (on a 9-point Likert scale with higher score indicating greater clinical benefit/importance).

We then refined the research questions as appropriate. We used the external stakeholders' ratings of clinical benefit/importance to classify each research question as follows:

- *high* clinical benefit/importance (between 7 and 9),
- *medium* clinical benefit/importance (between 4 and 6), and
- *low* clinical benefit/importance (between 1 and 3).

We defined consensus to have been achieved if at least 75 percent (7 out of 9) of external stakeholders rated clinical benefit/importance within a single category (high, medium, or low). Research questions for which consensus had been achieved were not retained in the list of research questions for the next Delphi round(s). The wording and clinical benefit/importance of these questions were deemed to have been agreed upon.

In the next two Delphi rounds, we provided stakeholders with the refined research questions for which consensus was not achieved. For each question, stakeholders were provided with a summary (mean, range) of the ratings of clinical benefit/importance those questions received in the previous Delphi round. In addition, we provided a brief synopsis of comments (with identifying information removed).

Step 6—Prioritization of Outcomes for Key Questions I and II

Because key questions I and II each had a long list of short- and long-term maternal and neonatal outcomes, we asked external stakeholders to rank the top three outcomes (in research questions related to those key questions) that would be essential to include in a clinical trial to address these research questions.

Step 7—Refinement of Final Research Questions and Development of Conceptual Models To Display Research Gaps

For each key question, we developed a conceptual model to identify and illustrate the location of research questions of high clinical benefit/importance and their relevant outcomes.

Step 8—Evaluation of Entire Process by Evidence Report Authors, Local Stakeholders, and External Stakeholders

We asked all those who participated in the project (other authors of the 2008 evidence report on GDM, local stakeholders, and external stakeholders) to evaluate our process. We used an online tool (SurveyGizmo,™ Widgix LLC, Boulder, CO) to obtain their feedback. Respondents were provided a summary of the process (Methods section of this report) and preliminary results (including tables, figures, and appendixes). They were asked to comment on:

- whether or not they felt they had had adequate information to participate effectively;
- which mode of participation they would have preferred (i.e., web-based form, phone, in-person);
- whether or not, after looking at the final list of research questions, they felt that we accomplished our objective;
- whether or not the local stakeholder group was comprehensive (i.e., whether it was adequate to only include physicians (obstetricians and gynecologists), nutritionists/dietitians, epidemiologists/methodologists, and patient/consumer representatives);
- whether or not the external stakeholder group was comprehensive (i.e., whether it was adequate to only include physicians (obstetricians and gynecologists), nutritionists/dietitians, epidemiologists/methodologists, research funders, and patient/consumer representatives);
- whether or not it was useful for us to get feedback from the report authors, local stakeholders, and external stakeholders; or if we could have abbreviated our process in some way; and
- any additional feedback.

Results

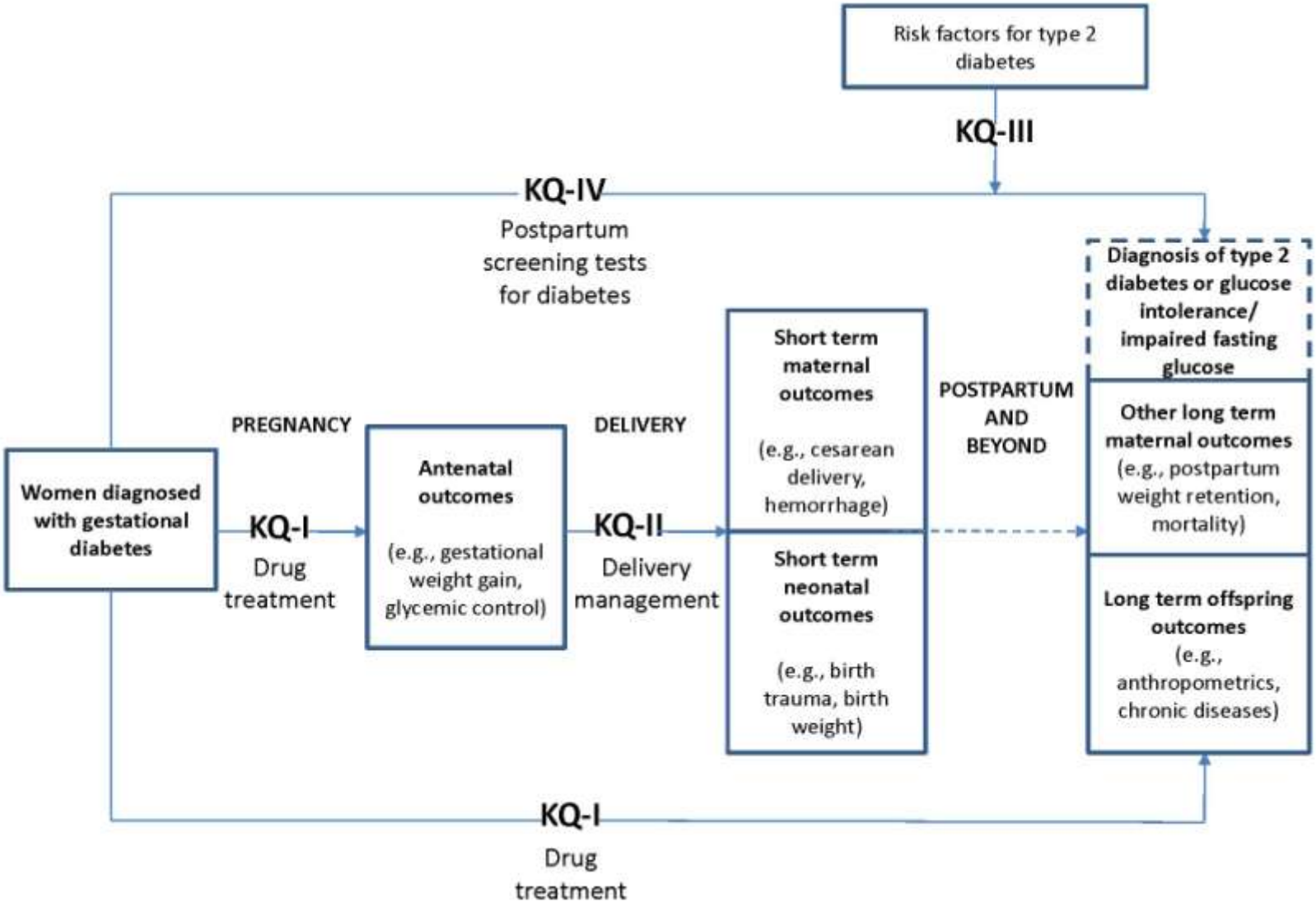
Step 1—Identification and Abstraction of Research Gaps from Evidence Report (2008) and Five Published Manuscripts

Our research team abstracted research gaps from the 2008 evidence report on GDM[7] and from the five published manuscripts[8-12] derived from the report (see Appendix A). We organized these research gaps by key question and within each key question, by gap area using the PICO framework. Concerns with limitations in study designs were also presented by key question. Specific gaps included research in diverse racial/ethnic populations and the type of GDM (diet-controlled, oral medications-requiring, insulin-requiring); the interventions and comparisons that had been the focus on the 2008 report; and many of the same short-term and long-term outcomes included in the 2008 report. In addition, we highlighted some of the challenges related to study design, such as the need for well-designed long-term observational cohort studies to better understand the progression to type 2 diabetes in women with GDM.

In addition to gaps that were specific to a key question, we abstracted general research gaps, also in the PICO and study design format. Examples included interventions to improve compliance with postpartum screening for type 2 diabetes and the comparative effectiveness of strategies (medication, behavioral) to prevent obesity and various forms of glucose intolerance (insulin resistance, impaired fasting glucose, impaired glucose tolerance, type 2 diabetes, etc.) in women with GDM (see Appendix A).

Figure 2 shows the conceptual model depicting the overall location of each of the original report's key questions and the main types of outcomes. Due to the widespread deficiencies in the literature, this conceptual model only outlines the overall location of the key questions and examples of types of interventions and outcomes. Specific research questions are outlined in individual conceptual models for each key question.

Figure 2. Overall conceptual model for all key questions (KQ-I to KQ-IV) from original (2008) evidence report on gestational diabetes



Step 2—Feedback from Authors of 2008 Evidence Report

Five out of eight of the other authors of the 2008 report (response rate=62.5 percent) provided feedback on the gaps identified in Step 1. We included all of their comments, clarifications, and additional suggested gaps in Appendix A.

For key question I (see Table 1), the authors highlighted the need for future studies to use consistent definitions of clinically relevant outcomes, such as hypoglycemia, to improve comparisons across studies. One author advised consideration of other medication comparators such as thiazolidinediones and dipeptidyl peptidase-4 (DPP-4) inhibitors. For key question II (see Table 1), authors stated that there is a need for randomized controlled trials (RCTs), but noting the challenges of conducting trials in this population, suggested that large prospective observational studies are needed to address this question. For key question III (see Table 1), authors suggested examining risk factors, such as psychosocial factors (e.g., depression, perceived stress), and their association with developing type 2 diabetes. For key question IV (see Table 1), authors suggested considering cluster randomized trials since they are more effectiveness-based, and may incorporate "real world" screening practices and outcomes.

Step 3—Translation of Research Gaps into Researchable Questions

We incorporated the feedback from the authors of the 2008 report and translated the list of gaps into a list of seventeen research questions (see Appendix B). Of these research questions, six related to key question I (benefits and harms of oral diabetes agents as compared to all types of insulin); three related to key question II (benefits and harms of elective cesarean delivery or the choice of timing of induction); five related to key question III (risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with GDM); and three related to key question IV (performance characteristics [sensitivity, specificity, and reproducibility] of tests for diagnosing type 2 diabetes following a pregnancy with GDM).

Step 4—Online and In-Person Feedback on Research Questions from Local Stakeholders

Online Feedback

We sent the seventeen research questions to the six local stakeholders for their online feedback. For each research question, we asked the stakeholders to provide specific comments on the clarity of wording, and rate (on a scale of 1=low to 9=high) the likely clinical benefit/importance and the likely ability to conduct a study that would address the research question. For each research question, Appendix B displays their comments, as well as the mean and range of their ratings of the clinical benefit/importance and the likely ability to conduct studies to address each research question. We received online feedback from each of the six local stakeholders (response rate=100 percent). The mean online feedback completion time was 78 minutes (range 20 to 289 minutes). We have summarized the results by key question below:

Key question I—benefits and harms of oral diabetes agents as compared to all types of insulin for treatment of GDM. We identified six research questions related to key question I.

Research question I-1 (comparative effectiveness and safety of any second generation sulfonylurea vs. any insulin) and research question I-2 (comparative effectiveness and safety of metformin vs. any insulin) received the highest ratings for clinical benefit/importance (each with mean 7.8, range 7 to 9). Research question I-3 (comparative effectiveness and safety of any oral hypoglycemic medication [i.e., a second generation sulfonylurea or metformin] vs. any insulin) was rated as the question researchers were most likely to be able to conduct a study to address (mean 7.8, range 7 to 9).

Key question II—benefits and harms of elective cesarean delivery or the choice of timing of induction. We identified three research questions related to key question II. Research question II-1 (comparative effectiveness and safety of elective labor induction vs. expectant management at term) received the highest rating for clinical benefit/importance (mean 7.7, range 7 to 8). Research question II-3 (comparative effectiveness and safety of elective labor induction or cesarean delivery vs. expectant management at term in women with insulin-requiring [class A2] GDM) received the highest rating for likely ability to conduct a study (mean 6.0, range 4 to 7).

Key question III—risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with GDM. We identified five research questions related to key question III. Research question III-1 (the effect of maternal lifestyle risk factors [e.g., behavioral risk factors, breastfeeding, physical activity, daily caloric intake]) received the highest rating for clinical benefit/importance (mean 8.2, range 7 to 9). Research question III-4 (the effect of maternal co-morbidities [e.g., obesity, hypertension, hypercholesterolemia]) received the highest rating for likely ability to conduct a study (mean 7.0, range 6 to 9).

Key question IV—performance characteristics (sensitivity, specificity, and reproducibility) of tests for diagnosing type 2 diabetes following a pregnancy with GDM. We identified three research questions related to key question IV. Research question IV-1 (accuracy of a single FBG test compared to the full 2-hour 75-gm OGTT and whether the comparative accuracy varies with the postpartum testing interval) received the highest ratings for clinical benefit/importance (mean 7.7, range 6 to 9) and likely ability to conduct a study (mean 6.8, range 5 to 8).

In-Person Meeting To Provide Additional Feedback

We revised the seventeen questions based on the local stakeholders' online feedback. We prepared a presentation to provide background and share the results of the 2008 report, and presented the revised questions and changes at the in-person followup meeting. Five of the six local stakeholders attended the meeting (response rate = 83.3 percent) and it lasted 90 minutes. For each research question, Appendix B displays the comments from the in-person meeting in the last column. We have summarized the comments by key question below:

Key question I—benefits and harms of oral diabetes agents as compared to all types of insulin for treatment of GDM. Overall, the stakeholders agreed that the six included research questions were clinically relevant. Some expressed concern about conducting studies of the newer medications, which have little evidence for use in pregnancy. There was agreement of the need to include patient-centered outcomes, such as quality of life, as well as long-term developmental and growth outcomes of the offspring.

Key question II—benefits and harms of elective cesarean delivery or the choice of timing of induction. Stakeholders discussed the use of the term "delivery" vs. "labor." We subsequently elected to change the questions to include the term "delivery," as well as to clarify that "term" is considered 40-weeks gestation. The stakeholders also highlighted the need for observational studies to include data with variation in both maternal and estimated fetal weight (EFW) to enable stratification. Stakeholders agreed that designing and conducting RCTs to address key question II would have many challenges, particularly with regards to providers not following random allocation and high levels of crossover between arms.

Key question III—risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with GDM. We asked stakeholders to review the list of risk factors included in the research questions in key question III. The stakeholders suggested inclusion of eating disorders and anxiety disorders among the psychosocial factors, interpregnancy interval, and number of pregnancies with GDM.

Key question IV—performance characteristics [sensitivity, specificity, and reproducibility] of tests for screening for type 2 diabetes following a pregnancy with GDM. Stakeholders agreed that patient and provider compliance with postpartum testing are probably of higher importance than assessment of the performance characteristics of the various screening tests.

The following questions were removed as separate questions, but the concepts from these questions were included in revised questions that were retained:

1. What is the comparative effectiveness and safety of any oral hypoglycemic medication (i.e., a second generation sulfonylurea or metformin) compared with any insulin in the treatment of gestational diabetes with regard to the following maternal and neonatal outcomes? (Key Question I)
2. What is the comparative effectiveness and safety of a short-acting insulin compared to diet alone in the treatment of gestational diabetes with regard to maternal and neonatal outcomes? (Key Question I)
3. What is the comparative effectiveness and safety of elective labor induction or cesarean delivery compared to expectant management at term in the management of labor in women with insulin-requiring (class A2) gestational diabetes with regard to the following maternal and neonatal outcomes? (Key Question II)
4. What is the reproducibility of the 2-hour 75-gram OGTT vs. a single fasting blood glucose test vs. a single HbA1c test in diagnosing type 2 diabetes following a pregnancy with gestational diabetes? (Key Question IV)

The following questions were added based on feedback from local stakeholders:

1. What is the evidence that the inter-conception interval is associated with the risk of developing type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes? (Key Question III)
2. What is the evidence that family history, gene mutations, genotypes, gene-environment interactions, epigenetic modifications, or other biomarkers are associated with the risk of developing type 2 diabetes among women with gestational diabetes? Are there differences in these associations by race or ethnic group? (Key Question III)

3. What is the comparative effectiveness of various lifestyle interventions for prevention of type 2 diabetes mellitus, glucose intolerance, and obesity in women with a history of gestational diabetes? (Key Question III)
4. What is the comparative effectiveness of various innovative strategies and technologies to disseminate educational materials on prevention of type 2 diabetes mellitus and glucose intolerance to women with a history of gestational diabetes? (Key Question III)
5. What is the comparative effectiveness of various strategies or interventions to improve clinician compliance with postpartum screening guidelines for type 2 diabetes mellitus and glucose intolerance in women with a history of gestational diabetes? (Key Question IV)
6. What is the comparative effectiveness of various innovative strategies or technologies (e.g., electronic health records) to track postpartum screening for the development of type 2 diabetes mellitus and glucose intolerance in women with a history of gestational diabetes? (Key Question IV)

Step 5—Online Feedback, Consensus Development, and Prioritization of Research Questions by External Stakeholders (Delphi Approach)

We further refined the research questions after feedback and suggestions from the local stakeholders at the in-person meeting. This refined list included nineteen research questions: four related to key question I, two related to key question II, nine related to key question III, and four related to key question IV. We sent these nineteen research questions to the nine external stakeholders for their feedback. The stakeholders participated in three Delphi rounds, with the goal of providing constructive feedback about the wording and clarity of each of the research questions and developing consensus on the likely clinical benefit/importance of addressing the research question. Stakeholders rated the latter question on a 9-point Likert scale (1=lowest clinical benefit/importance and 9=highest clinical benefit/importance).

We achieved a response rate of 100 percent for each Delphi round. We have described below the results from each of the Delphi rounds and highlight the sixteen research questions that achieved consensus by the third round (rated as low, medium, or high clinical benefit/importance) and the three that did not achieve consensus by the third round.

Delphi Round 1

In Delphi round 1, we sent the external stakeholders each of the nineteen research questions (see Appendix C). The mean online feedback completion time for Delphi round 1 by external stakeholders was 63 minutes (range 21 to 112 minutes). Among these nineteen research questions, consensus of high clinical benefit/importance was established on eight (42.1 percent) questions.

Two (I-1 and I-2) of four research questions related to key question I, two (II-1 and II-2) of two research questions related to key question II, three (III-1, III-7, and III-8) of nine questions related to key question III, and one (IV-3) of four questions related to key question IV achieved consensus of having high clinical benefit/importance (see Appendix C).

Regarding the two research questions (I-1 and I-2) related to key question I that were rated as high clinical benefit/importance, stakeholders expressed concern that use of oral hypoglycemic agents is now common clinical practice even though their effectiveness and safety

has not clearly been established (few published RCTs). There was also interest in examining long-term effects of treatment on offspring, particularly metformin as an insulin-sensitizer. Research question I-3 did not achieve consensus because there was confusion about how the insulin comparisons were phrased. This was clarified based on feedback from the stakeholders. For research question I-4 stakeholders expressed concerns about safety of use of hypoglycemic medications newly approved by the Food and Drug Administration (FDA), both in and outside of pregnancy (see Appendix D).

The two research questions (II-1 and II-2) related to key question II were considered of high clinical benefit/importance because of a dearth of evidence in this clinically important area. Stakeholders emphasized the importance of including patient-oriented outcomes for these questions.

For key question III, stakeholders agreed that the overall question identifying risk factors for type 2 diabetes in women with GDM was of high importance, particularly to guide future preventive interventions. In this Delphi round, they highlighted which risk factors were of the highest clinical importance for future research: maternal lifestyle factors (III-1); family history, gene mutations, genotype and gene-environment interactions (III-7); and lifestyle interventions (III-8).

For key question IV, stakeholders agreed that focusing on low compliance with postpartum screening for type 2 diabetes (IV-3) was more clinically important than the other questions addressing the screening tests' performance characteristics (see Appendix D).

External stakeholders' comments about study needs and challenges. In Delphi round 1, in addition to comments about clinical importance, the external stakeholders commented on the study needs and challenges identified during the in-person local stakeholder meeting. These comments are included in Appendix D. For key question I, the external stakeholders emphasized the importance of having a standard GDM diagnosis; designing studies with longer term followup of offspring to assess obesity; and suggested the inclusion of a research question addressing optimal glucose thresholds for starting treatment. For key question II, external stakeholders agreed with the local stakeholders on the practical challenges and barriers to designing an RCT to address these research questions, in particular in women without evidence of elevated EFW. For key question III, stakeholders advised including the intervention of lactation support. They also emphasized the importance of when risk factors were measured—pre, during, and/or post pregnancy (this was subsequently clarified for relevant questions). For key question IV, stakeholders suggested that the ideal study design to assess performance characteristics of screening tests for diabetes would include long-term assessment of complications.

Delphi Round 2

In Delphi round 2, we sent the external stakeholders the eleven research questions that had not achieved consensus in Delphi round 1. The mean online feedback completion time for Delphi round 2 by external stakeholders was 63 minutes (range 11 to 355 minutes). Consensus was achieved on six (54.6 percent) of these eleven questions.

Consensus of high clinical benefit/importance was established on five (45.5 percent) research questions. These included one (I-3) of two questions related to key question I, three (III-3, III-4, and III-9) of six questions related to key question III, and one (IV-2) of three questions

related to key question IV. Consensus of medium clinical benefit/importance was established on one research question (III-6) related to key question III (see Appendix C).

After re-wording research question I-3 for clarity of the insulin comparisons, consensus of high clinical benefit/importance was established in Delphi round 2. For research question III-4, we re-structured the question to specify the time periods for risk factor measurement as "prepregnancy," "antenatal," and "postpartum." The question then received consensus of high clinical benefit/importance. During all three Delphi rounds, only one research question achieved consensus of medium clinical benefit/importance. This was research question III-6, focused on inter-conception interval as a risk factor. Stakeholders highlighted the high degree of confounding with this question, such as access to care, contraception, and postpartum weight retention. They thus did not rate it as highly as some of the other risk factors. For key question IV, stakeholders had higher interest in comparing HbA1c as a postpartum screening test (IV-2) than FBG (IV-1) with the gold standard OGTT (see Appendix E).

Delphi Round 3

In Delphi round 3, we sent the external stakeholders the five research questions that had not achieved consensus in Delphi round 2. The mean online feedback completion time for Delphi round 3 by external stakeholders was 16 minutes (range 3 to 31 minutes). Among these five research questions, consensus of high clinical benefit/importance was achieved on two (40 percent) questions. These questions included one research question (I-4) related to key question I and one (IV-1) of two questions related to key question IV (see Appendix C).

Stakeholders achieved consensus on research question I-4, related to the use of newly FDA-approved oral hypoglycemic medications to treat GDM. Although consensus was reached, two stakeholders emphasized the importance of establishing the safety of metformin and glyburide before beginning research on newer agents not currently used in clinical practice. For key question IV, stakeholders reached consensus to include research question IV-1 (comparing a single FBG with the OGTT) as one of high clinical benefit/importance (see Appendix F).

Research Questions for Which No Consensus Was Established After Three Delphi Rounds

Table 3 lists the research questions for which no consensus was established. These included two research questions (III-2 and III-5) related to key question III and one (IV-4) related to key question IV.

Table 3. Research questions where no consensus was established or consensus of medium or low clinical benefit/importance was established

Research Question Number	Research Questions	Consensus Status*
III-6	What is the evidence that the inter-conception interval is associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Consensus - MEDIUM IMPORTANCE
III-2	What is the evidence that maternal psychosocial factors (e.g., mood disorders, substance use disorders, eating disorders, stress) are associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	No consensus
III-5	What is the evidence that contraceptive method (e.g., progestin-only) is associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	No consensus
IV-4	What is the comparative effectiveness of health information technology interventions to track postpartum screening for the development of type 2 diabetes and glucose intolerance/impaired fasting glucose in women with a history of gestational diabetes?	No consensus

*We used the external stakeholders' ratings of clinical benefit/importance to classify each research question as follows:
• *high* clinical benefit/importance (between 7 and 9),
• *medium* clinical benefit/importance (between 4 and 6), and
• *low* clinical benefit/importance (between 1 and 3).
We defined consensus to have been achieved if at least 75 percent (7 out of 9) of stakeholders rated clinical benefit/importance within a single category (high, medium, or low).

Research question III-2 addressed the association of maternal psychosocial factors (e.g., mood disorders, substance use disorders, eating disorders, stress) with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with GDM. Stakeholders expressed uncertainty about the relevance of these risk factors to pregnancy, compared to the development of diabetes in the general population. Research question III-5 addressed whether contraceptive method (e.g., progestin-only) was associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with GDM. Some stakeholders determined this question to be of lower clinical importance. This question was complicated by timing of contraceptive use and consistency/compliance with a method. Some stakeholders expressed that the impact of progestin-only contraceptive methods on glucose tolerance was known and that the research question did not address a knowledge gap.

One research question related to key question IV did not achieve consensus. Research question IV-4 addressed the comparative effectiveness of health information technology interventions to track postpartum screening for the development of type 2 diabetes mellitus and glucose intolerance/impaired fasting glucose in women with a history of GDM. This was a newer question that had been added during the in-person discussion with local stakeholders. However, there was no consensus among external stakeholders about the clinical benefit/importance of this question.

Research Question for Which Consensus of Medium Clinical Benefit/Importance Was Established

Table 3 also lists the one research question (III-6) for which consensus of medium clinical benefit/importance was established. This question, related to key question III, addressed

whether inter-conception interval is associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with GDM.

Step 6—Prioritization of Outcomes for Key Questions I and II

Prioritized Outcomes for Research Questions Related to Key Question I

Research questions related to key question I included thirty outcomes of interest (thirteen short-term maternal outcomes, four long-term maternal outcomes, ten neonatal outcomes, and three long-term offspring outcomes). Table 4 lists the prioritized outcomes for research questions related key question I (benefits and harms of oral diabetes agents as compared to all types of insulin). Four of nine stakeholders ranked the long-term offspring outcome of chronic diseases (e.g., obesity, type 2 diabetes) in one of their top three, making it the highest rated outcome. The next most highly rated outcomes were the short-term maternal outcomes of hypertensive disorders of pregnancy (e.g., gestational hypertension, preeclampsia) and medication adherence, and the neonatal outcome of large for gestational age and macrosomia. Three of nine stakeholders rated each of these three outcomes as one of their top three to study.

Table 4. Results from step 6 (prioritization of outcomes for research questions related to key question I (KQ-I))

Number (out of 9) of External Stakeholders who Ranked the Outcome as One of Their Top Three	Type of Outcome	Outcome
4	Long-term offspring	Chronic diseases (e.g., obesity, type 2 diabetes)
3	Short-term maternal	Hypertensive disorders of pregnancy (e.g., gestational hypertension, preeclampsia)
3	Short-term maternal	Medication adherence
3	Neonatal	Large for gestational age and macrosomia
2	Short-term maternal	Gestational weight gain
2	Short-term maternal	Hypoglycemia
2	Long-term maternal	Postpartum type 2 diabetes mellitus or glucose intolerance/impaired fasting glucose
2	Neonatal	Neonatal intensive care unit admission
1	Short-term maternal	Patient-reported outcomes (e.g., patient treatment preference, quality of life)
1	Short-term maternal	Shoulder dystocia
1	Short-term maternal	Glycemic control (e.g., fasting blood glucose, 2-hr postprandial glucose)
1	Neonatal	Hypoxia/anoxia
1	Neonatal	Birth trauma (e.g., bone fractures, brachial plexus palsy)
1	Neonatal	Birth weight

The following outcomes were not ranked in the top 3 by any external stakeholder:
Short-Term Maternal Outcomes—cesarean delivery (including primary cesarean and repeat cesarean) and indication for cesarean delivery, complications of cesarean delivery (e.g., wound infection, wound dehiscence), vaginal delivery (and specify type: spontaneous or operative), perineal lacerations, postpartum hemorrhage, and peripartum mortality
Long-Term Maternal Outcomes—postpartum weight retention, patient-reported outcomes (e.g., quality of life), and mortality
Neonatal Outcomes—hyperbilirubinemia, hypoglycemia, small for gestational age, respiratory distress syndrome, and perinatal mortality
Long-Term Offspring Outcomes—infant and child growth, and anthropometrics.

Prioritized Outcomes for Research Questions Related to Key Question II

Research questions related to key question II included nineteen outcomes of interest (nine maternal outcomes and ten neonatal outcomes). Table 5 lists the prioritized outcomes for research questions related key question II (benefits and harms of elective cesarean delivery or the choice of timing of induction). The most important outcome was the maternal outcome of cesarean delivery (including primary cesarean and repeat cesarean) and indication for cesarean delivery, which six of nine stakeholders ranked as one of their top three. The next most important outcomes were the neonatal outcomes of birth trauma (e.g., bone fractures, brachial plexus palsy), which four of nine stakeholders ranked as one of their top three, and neonatal intensive care unit (NICU) admission, which three of nine stakeholders ranked as one of their top three.

Table 5. Results from step 6 (prioritization of outcomes for research questions related to key question II (KQ-II)

Number (out of 9) of External Stakeholders who Ranked the Outcome as One of Their Top Three	Type of Outcome	Outcome
6	Maternal	Cesarean delivery (including primary cesarean and repeat cesarean) and indication for cesarean delivery
4	Neonatal	Birth trauma (e.g., bone fractures, brachial plexus palsy)
3	Neonatal	Neonatal intensive care unit admission
2	Maternal	Patient-reported outcomes (e.g., patient preference, quality of life)
2	Maternal	Complications of cesarean delivery (e.g., wound infection, wound dehiscence)
2	Maternal	Vaginal delivery (spontaneous, operative)
2	Neonatal	Hypoxia/anoxia
2	Neonatal	Respiratory distress syndrome
1	Maternal	Resource utilization (e.g., cost of care, length of hospital stay)
1	Maternal	Peripartum mortality
1	Neonatal	Birth weight
1	Neonatal	Hypoglycemia
The following outcomes were not ranked in the top 3 by any external stakeholder: **Maternal Outcomes**—perineal lacerations, postpartum hemorrhage, and pulmonary embolism **Neonatal Outcomes**—hyperbilirubinemia, large for gestational age and macrosomia, small for gestational age, and perinatal mortality.		

Step 7—Refinement of Final Research Questions and Development of Conceptual Models to Display Research Gaps

Final Research Questions

Through the three Delphi rounds, fifteen of the nineteen research questions achieved consensus of high clinical benefit/importance and one research question achieved consensus of medium clinical benefit/importance, yielding an overall Delphi consensus establishment rate of 84.2 percent.

Research questions of high clinical benefit/importance. Table 6 lists the fifteen final research questions of high clinical benefit/importance. These include four research questions related to key question I, two related to key question II, six related to key question III, and three related to

key question IV. The research question that was rated the highest (mean 8.2 on a scale of 1 to 9, range 7 to 9) in terms of clinical benefit/importance was research question I-1. This research question compared the effectiveness and safety of any of the second generation sulfonylureas with any insulin in the treatment of GDM. Research question III-1 was rated the next highest (mean 8.1 on a scale of 1 to 9, range 7 to 9) and addressed whether maternal health behaviors (e.g., breastfeeding, physical activity, diet) are associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with GDM.

Table 6. Final list of research questions rated as high clinical benefit/importance

Sr. No.	Question Number	Final Research Questions	Clinical Benefit/ Importance*** (Mean, Range on a Scale of 1-9)
1	I-1	What are the effectiveness and safety of any of the second generation sulfonylureas compared to any insulin in the treatment of gestational diabetes with regard to the following short- and long-term maternal outcomes, neonatal outcomes, and long-term offspring outcomes?*	Mean=8.2 Range=7 to 9
2	I-2	What are the effectiveness and safety of metformin compared to any insulin in the treatment of gestational diabetes with regard to the following short- and long-term maternal outcomes, neonatal outcomes, and long-term offspring outcomes?*	Mean=7.9 Range=6 to 9
3	I-3	What are the comparative effectiveness and safety of various insulin regimens in terms of type/duration, dosing, and frequency of administration in the treatment of gestational diabetes with regard to the following short- and long-term maternal outcomes, neonatal outcomes, and long-term offspring outcomes?*	Mean=7.3 Range=6 to 9
4	I-4	What are the effectiveness and safety of other hypoglycemic drug classes (e.g., thiazolidinediones, DPP-4 inhibitors, GLP-1 agonists, meglitinides) compared to any insulin or other hypoglycemic drugs in the treatment of gestational diabetes with regard to the following short- and long-term maternal outcomes, neonatal outcomes, and long-term offspring outcomes?*	Mean=6.9 Range=4 to 9
5	II-1	What are the effectiveness and safety of elective labor induction at 40 weeks compared to expectant management in women with gestational diabetes with regard to the following maternal and neonatal outcomes?** Populations of Interest: • All women with gestational diabetes • Women with insulin-requiring (class A2) gestational diabetes • Obese women with gestational diabetes • Women with gestational diabetes with high estimated fetal weight (e.g., >4000 or >4500 gram) • Women with different parities • Women of various races/ethnicities.	Mean=7.8 Range=6 to 9
6	II-2	What are the effectiveness and safety of elective cesarean delivery at 40 weeks compared to expectant management in women with gestational diabetes with regard to the following maternal and neonatal outcomes?** Populations of Interest: • All women with gestational diabetes • Women with insulin-requiring (class A2) gestational diabetes • Obese women with gestational diabetes • Women with gestational diabetes with high estimated fetal weight (e.g., >4000 or >4500 gram) • Women with different parities • Women of various races/ethnicities.	Mean=7.3 Range=4 to 9

Table 6. Final list of research questions rated as high clinical benefit/importance (continued)

Sr. No.	Question Number	Final Research Questions	Clinical Benefit/ Importance*** (Mean, Range on a Scale of 1-9)
7	III-1	What is the evidence that maternal health behaviors (e.g., breastfeeding, physical activity, diet) are associated with the risk of developing type 2 diabetes or glucose intolerance/ impaired fasting glucose following a pregnancy with gestational diabetes?	Mean=8.1 Range=7 to 9
8	III-3	What is the evidence that maternal metabolic measures (e.g., fasting insulin levels, OGTT measures, HPA axis stress (subclinical hypercortisolism)) are associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Mean=7.3 Range=6 to 8
9	III-4	What is the evidence that co-morbid conditions (e.g., advanced maternal age, obesity, hypertension, hypercholesterolemia) are associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Mean=7.4 Range=6 to 9
10	III-7	What is the evidence that family history, gene mutations, genotypes, gene-environment interactions, epigenetic modifications, or other biomarkers are associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose among women with gestational diabetes? Are there differences in these associations by race or ethnic group?	Mean=7.4 Range=3 to 9
11	III-8	What is the comparative effectiveness of various lifestyle interventions (e.g., diet, physical activity, smoking) for prevention of type 2 diabetes, glucose intolerance/impaired fasting glucose, and obesity in women with a history of gestational diabetes?	Mean=7.7 Range=6 to 9
12	III-9	What is the comparative effectiveness of various educational and behavioral change strategies (e.g., patient education about diabetes risk, lactation support, diet, physical activity) for prevention of type 2 diabetes and glucose intolerance/impaired fasting glucose in women with a history of gestational diabetes?	Mean=7.3 Range=2 to 9
13	IV-1	What are the performance characteristics (sensitivity, specificity, and reproducibility) of a single fasting blood glucose test compared to the full 2-hour 75-gm OGTT in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes? Does the accuracy of the fasting blood glucose test compared to the full 2-hour 75-gm OGTT vary with the postpartum testing interval in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Mean=6.7 Range=1 to 9
14	IV-2	What are the performance characteristics (sensitivity, specificity, and reproducibility) of the HbA1c test compared to the 2-hour 75-gm OGTT in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes? Does the accuracy of the HbA1c test compared to the full 2-hour 75-gm OGTT vary with the postpartum testing interval in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Mean=7.9 Range=6 to 9
15	IV-3	What is the comparative effectiveness of strategies or interventions to improve clinician compliance with postpartum screening guidelines for type 2 diabetes and glucose intolerance/impaired fasting glucose in women with a history of gestational diabetes?	Mean=7.8 Range=5 to 9

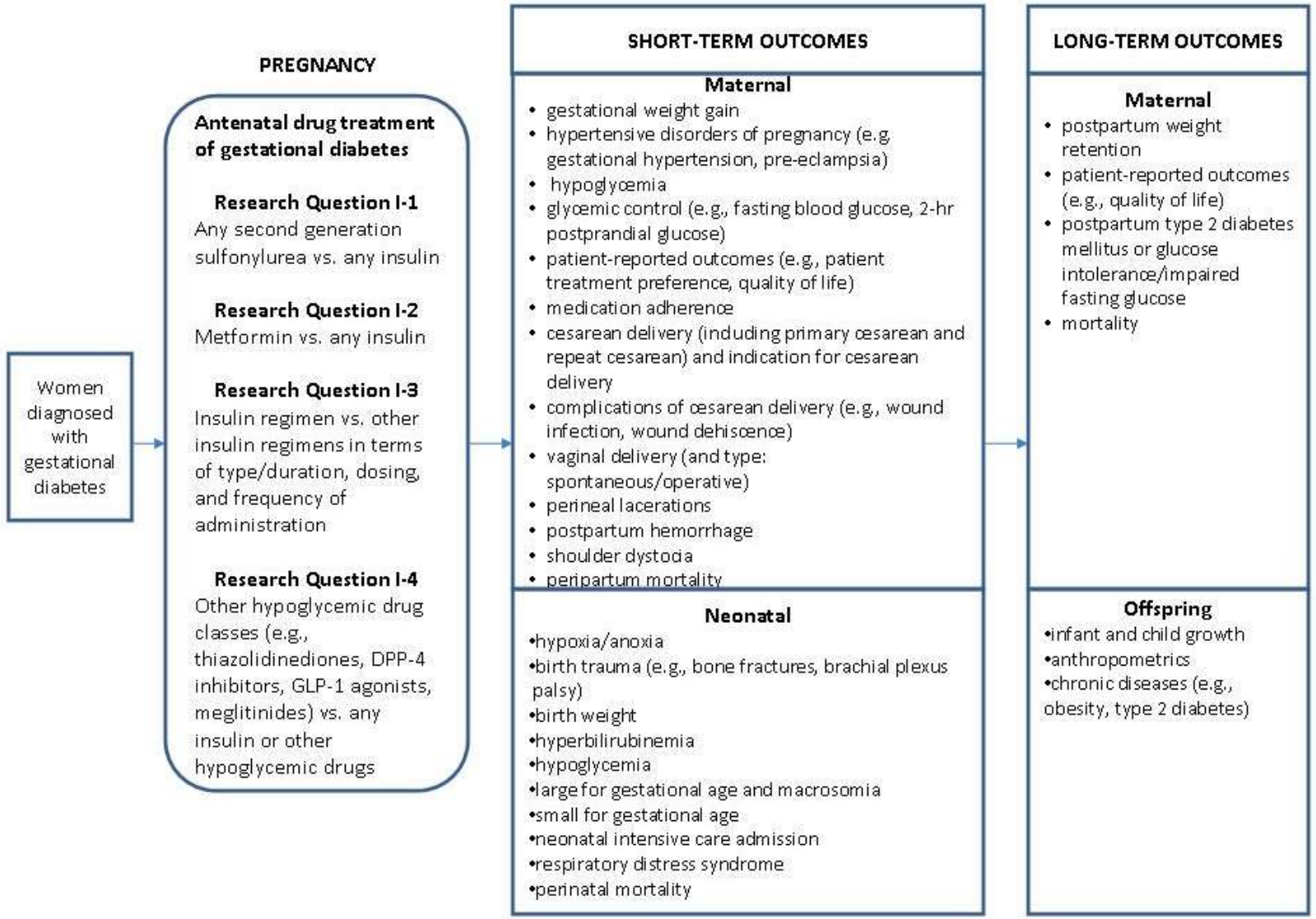
Table 6. Final list of research questions rated as high clinical benefit/importance (continued)

* Outcomes for Research Questions I-1, I-2, I-3, and I-4: **Short-Term Maternal Outcomes:** gestational weight gain, hypertensive disorders of pregnancy (e.g., gestational hypertension, preeclampsia), hypoglycemia, glycemic control (e.g., fasting blood glucose, 2-hr postprandial glucose), patient-reported outcomes (e.g., patient treatment preference, quality of life), medication adherence, cesarean delivery (including primary cesarean and repeat cesarean) and indication for cesarean delivery, complications of cesarean delivery (e.g., wound infection, wound dehiscence), vaginal delivery (and specify type: spontaneous or operative), perineal lacerations, postpartum hemorrhage, shoulder dystocia, and peripartum mortality **Long-Term Maternal Outcomes:** postpartum weight retention, obesity, patient-reported outcomes (e.g., quality of life), development of postpartum type 2 diabetes or glucose intolerance/impaired fasting glucose, and mortality **Neonatal Outcomes:** hypoxia/anoxia, birth trauma (e.g., bone fractures, brachial plexus palsy), birth weight, hyperbilirubinemia, hypoglycemia, large for gestational age and macrosomia, small for gestational age, neonatal intensive care admission, respiratory distress syndrome, and perinatal mortality **Long-Term Offspring Outcomes:** infant and child growth, anthropometrics, and chronic diseases (e.g., obesity, type 2 diabetes).
** Outcomes for Research Questions II-1 and I-2: Maternal Outcomes: cesarean delivery (including primary cesarean and repeat cesarean) and indication for cesarean delivery, complications of cesarean delivery (e.g., wound infection, wound dehiscence), vaginal delivery (spontaneous, operative), perineal lacerations, hemorrhage, patient-reported outcomes (e.g., patient preference , quality of life), length of hospital stay, pulmonary embolism, and mortality Neonatal Outcomes: hypoxia/anoxia, birth trauma (e.g., bone fractures, brachial plexus palsy), birth weight, hyperbilirubinemia, hypoglycemia, large for gestational age and macrosomia, small for gestational age, neonatal intensive care admission, respiratory distress syndrome, and perinatal mortality.
***We used the external stakeholders' ratings of clinical benefit/importance to classify each research question as follows: • "high" clinical benefit/importance (between 7 and 9), • "medium" clinical benefit/importance (between 4 and 6), and • "low" clinical benefit/importance (between 1 and 3). We defined consensus to have been achieved if at least 75 percent (7 out of 9) of stakeholders rated clinical benefit/importance within a single category (high, medium, or low).
Abbreviations: DPP-4=dipeptidyl peptidase-4, GLP-1=glucagon-like peptide-1, HbA1c=hemoglobin A1c, HPA=hypothalamo-pituitary-adrenal, OGTT=oral glucose tolerance test.

Development of Conceptual Models To Display Research Gaps

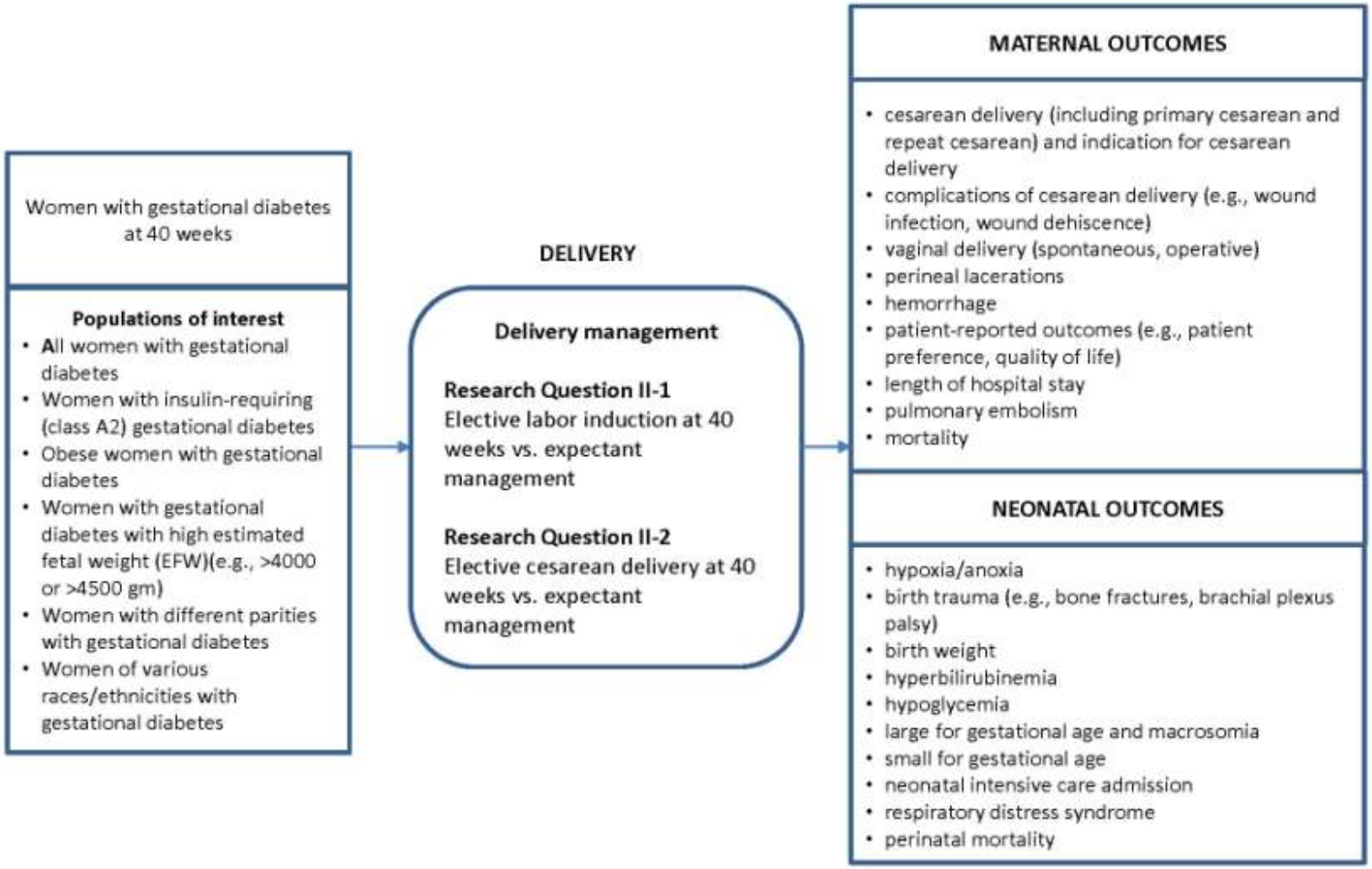
For each of the original key questions, we developed a conceptual model to pictorially display the identified research questions, according to the population, and outcomes (see Figures 3, 4, 5, and 6 for conceptual models for research questions related to key questions I, II, III, and IV respectively).

Figure 3. Results from step 7 (conceptual model to display research questions related to key question I (KQ-I))



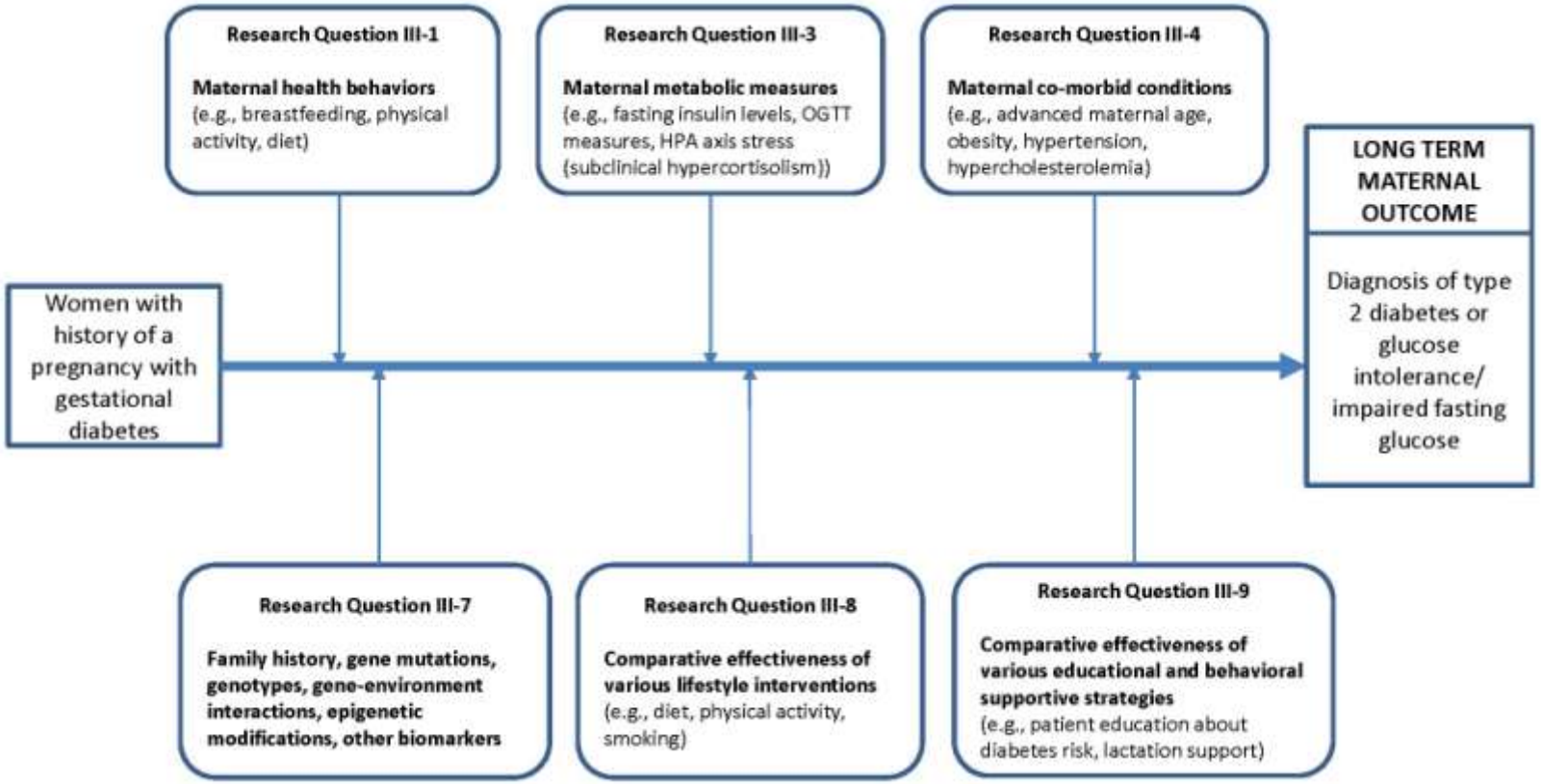
Abbreviations: DPP-4=dipeptidyl peptides-4, GLP-1=glucagon-like peptide-1.

Figure 4. Results from step 7 (conceptual model to display research questions to key question II (KQ-II))



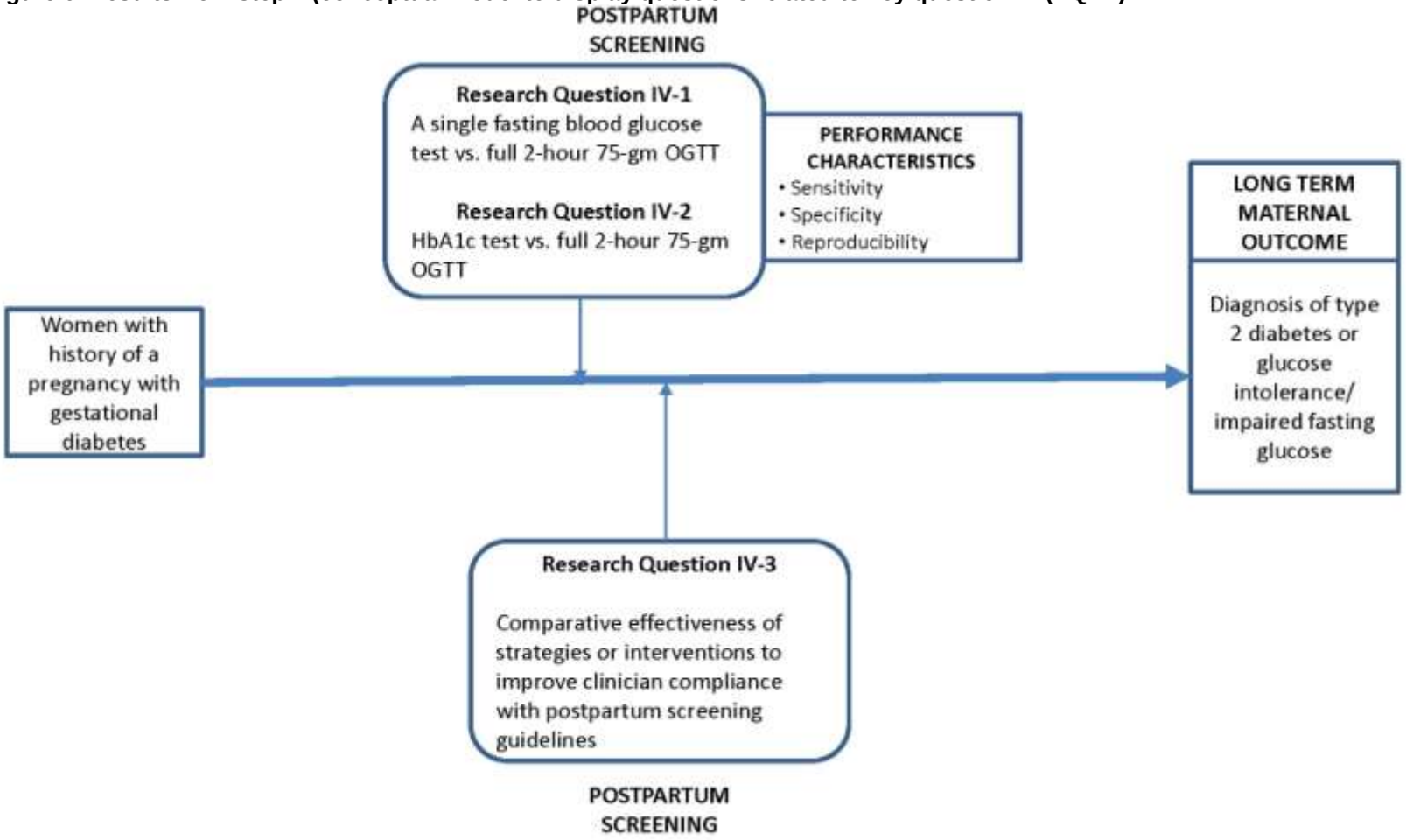
Abbreviations: EFW=estimated fetal weight.

Figure 5. Results from step 7 (conceptual model to display research questions related to key question III (KQ-III)



Abbreviations: HPA=hypothalamo-pituitary-adrenal, OGTT=oral glucose tolerance test.

Figure 6. Results from step 7 (conceptual model to display questions related to key question IV (KQ-IV)



Abbreviations: HbA1c=hemoglobin A1c, OGTT=oral glucose tolerance test.

Step 8—Evaluation of Entire Process by Evidence Report Authors, Local Stakeholders, and External Stakeholders

We sent an online evaluation form to the twenty contributors to this project, which included five authors of the 2008 evidence report, six local stakeholders, and nine external stakeholders. We received feedback from each of the twenty contributors (response rate=100 percent). The mean feedback completion time for the evaluation form was 12 minutes (range 2 to 72 minutes). Contributors were asked to review the report materials (text, tables, figures, and appendixes) before completing the form. The time taken to review these materials was not recorded.

Responses to the evaluation are listed in Appendix G for the 2008 report authors and the local stakeholders and in Appendix H for the external stakeholders. All twenty contributors felt that they had adequate information to effectively participate and that we had accomplished our objective of identifying important research questions for GDM. A web-based form was the most preferred mode of participation (fifteen contributors, 75 percent). Two (10 percent) contributors (both local stakeholders) preferred in-person participation to a web-based form, while another two (one 2008 evidence report author and one local stakeholder) preferred a combination of a web-based form and in-person participation. One (5 percent) contributor preferred participation via telephone.

Fifteen (79 percent) of nineteen contributors felt that the composition of the local group of stakeholders was comprehensive. One external stakeholder felt unable to comment on the comprehensiveness of the local stakeholder group. Fifteen (75 percent) of twenty contributors felt that the composition of the external group of stakeholders was comprehensive. The following additional perspectives were suggested—endocrinologists/diabetologists managing women with GDM, neonatologists, nephrologists, and patients with current GDM or with a history of GDM.

One (5 percent) contributor felt that we could have abbreviated our process, questioning the need for a local stakeholder group.

Discussion

Using the 2008 JHU EPC evidence report on the management and postpartum followup of GDM, we developed an eight-step process for identifying and prioritizing clinically important research needs, with key input from a diverse group of stakeholders. The research needs, reflecting the breadth of the original key questions, address a variety of interventions, risk factors for development of GDM and outcomes. Questions that were specifically added through the involvement with stakeholders included the role of genetics in the development of GDM and the role of lifestyle changes in prevention, as well as questions about appropriate ways to increase patient and physician education and compliance. There did not appear to be a particular type of question that did not reach consensus or was not rated as of high clinical importance. Through this process we propose a final list of fifteen research questions, and high priority outcomes of interest, which highlight the most up-to-date research needs in the field.

There are several strengths to the process we developed. First, our research team had diverse expertise. Three of the members of the team were also part of the original report's research team. In fact, one of the co-principal investigators of this project (WKN) was the principal investigator of the original report. Their insight into the report and the field of GDM enabled greater depth into the field and in the selection of stakeholders. Second, we invited stakeholders from a wide range of relevant disciplines, including social work and nutrition, to ensure a balanced and broad perspective on research needs for GDM. Each stakeholder was highly interested and committed. There were high levels of participation at each step. Third, we used the Delphi method to achieve formal consensus development.

There are several limitations to our process for identifying research gaps. In particular, we had limited input from patients and patient liaisons. We communicated with non-research nutritionists and clinicians in the local stakeholder step, but their feedback was limited by the complexity of the project and amount of information presented from the 2008 report. In fact, the social worker (a local stakeholder) proposed examining psychosocial risk factors for progression to type 2 diabetes, but this was ranked as lower priority in the Delphi rounds with external stakeholders. In addition, this was a fairly resource-intense process, with eight steps, including three Delphi rounds. It is possible that the same research team that just completed an evidence report may not be able to commit the time and resources to these steps. Rather, we would suggest an independent or combination independent-current evidence report team carry out this additional process.

Finally, we developed and completed eight steps to identify and prioritize research needs. A research or evidence gap is a topic or area for which missing or inadequate information limits the ability of reviewers to reach a conclusion on a given question. Steps 1 through 3 identified research gaps and translated these to research questions. The remaining steps identified research needs. A research need is a topic or area for which further research is needed to fill a research gap. A research gap may not be a research need if filling the gap does not help decisionmakers. There are additional steps to be completed in taking these research needs and developing a research agenda. Additional steps would include the search for new evidence that may fill the identified evidence gaps, such as an update of the search from the 2008 evidence report and a search for ongoing studies. Also needed for the development of a research agenda are decisions about appropriate study designs and the feasibility of addressing the identified research needs given local circumstances such as funding and current practice. These additional steps would involve a different process, including the participation of different stakeholders.

Lessons Learned and Future Directions

Our team greatly learned from this process and it is an opportunity to share lessons learned. Although stakeholders generally provided positive feedback about completing web-based materials, several agreed that more interactive approaches, such as Webinars would have been more effective in engaging multiple stakeholders from various geographical regions. We also identified the need to involve a large number of non-research-oriented clinician and patient perspectives in the process, especially related to studying patient-oriented outcomes. For this pilot we recognized certain logistical barriers to involving patients, such as institutional review board (IRB) review, selection of a representative sample of patients, and the conduct of qualitative interviews. An alternative strategy for future research gaps identification projects might be to include former patients referred or recommended by advocacy organizations, such as but not limited to the American Diabetes Association (ADA), American College of Obstetricians and Gynecologists (ACOG), and the March of Dimes. Future projects will need to incorporate patients and their perspectives, particularly for prioritizing patient-centered outcomes of interest. One stakeholder questioned whether we needed all eight steps to reach the same conclusions. Our research team determined that each step was useful, often iterative, and built upon the previous step. Moreover, each step helped to ensure a broad, comprehensive, and unbiased summary of current research gaps. However, it may be that certain steps could be abbreviated. The impact of any changes should be evaluated.

Conclusions

Using the 2008 evidence report as a starting point, we developed an eight-step process and identified fifteen research questions on GDM considered of high clinical importance by a multidisciplinary group of stakeholders. We also prioritized outcomes of highest clinical benefit related to six of these questions. In addition, we evaluated the process through a feedback mechanism from the evidence report authors and local and external stakeholders. We anticipate that our process could be used as is, or modified based on the evaluations, as a model for using other evidence reports to take the next step of developing research questions in areas of highest clinical importance. This next step would enable researchers and funding agencies to focus their resources in areas of highest need to make the most clinical impact.

References

1. Metzger BE, Coustan DR. Summary and recommendations of the Fourth International Workshop-Conference on Gestational Diabetes Mellitus. The Organizing Committee. Diabetes Care 1998;21(Suppl 2):B161-B167.
2. Beckles GL, Thompson-Reid PE, Eds. Diabetes and women's health across the life stages: A public health perspective. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Diabetes Translation, 2001.
3. Hunt KJ, Schuller KL. The increasing prevalence of diabetes in pregnancy. Obstet Gynecol Clin North Am 2007;34(2):173-199, vii.
4. Ferrara A, Hedderson MM, Quesenberry CP, et al. Prevalence of gestational diabetes mellitus detected by the national diabetes data group or the carpenter and coustan plasma glucose thresholds. Diabetes Care 2002;25(9):1625-1630.
5. Rosenberg TJ, Garbers S, Lipkind H, et al. Maternal obesity and diabetes as risk factors for adverse pregnancy outcomes: differences among 4 racial/ethnic groups. Am J Public Health 2005;95(9):1545-1551.
6. Feig DS, Zinman B, Wang X, et al. Risk of development of diabetes mellitus after diagnosis of gestational diabetes. CMAJ 2008;179(3):229-234.
7. Nicholson WK, Wilson LM, Witkop CT, et al. Therapeutic management, delivery, and postpartum risk assessment and screening in gestational diabetes. Evid Rep Technol Assess (Full Rep) 2008;(162):1-96.
8. Bennett WL, Bolen S, Wilson LM, et al. Performance characteristics of postpartum screening tests for type 2 diabetes mellitus in women with a history of gestational diabetes mellitus: a systematic review. J Womens Health (Larchmt) 2009;18(7):979-987.
9. Golden SH, Bennett WL, Baptist-Roberts K, et al. Antepartum glucose tolerance test results as predictors of type 2 diabetes mellitus in women with a history of gestational diabetes mellitus: a systematic review. Gend Med 2009;6(Suppl 1):109-122.
10. Baptiste-Roberts K, Barone BB, Gary TL, et al. Risk factors for type 2 diabetes among women with gestational diabetes: a systematic review. Am J Med 2009;122(3):207-214.e4.
11. Witkop CT, Neale D, Wilson LM, et al. Active compared with expectant delivery management in women with gestational diabetes: a systematic review. Obstet Gynecol 2009;113(1):206-217.
12. Nicholson W, Bolen S, Witkop CT, et al. Benefits and risks of oral diabetes agents compared with insulin in women with gestational diabetes: a systematic review. Obstet Gynecol 2009;113(1):193-205.
13. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ 1995;311(7001):376-380.

Abbreviations

ACOG	American College of Obstetricians and Gynecologists
ADA	American Diabetes Association
AHRQ	Agency for Healthcare Research and Quality
BMI	Body Mass Index
DPP-4	Dipeptidyl Peptidase-4
EFW	Estimated Fetal Weight
FBG	Fasting Blood Glucose
FDA	Food and Drug Administration
GDM	Gestational Diabetes Mellitus
GLP-1	Glucagon-Like Peptide-1
HbA1c	Hemoglobin A1c
HDL	High Density Lipoproteins
HPA	Hypothalamic-Pituitary-Adrenal
IRB	Institutional Review Board
JHU EPC	Johns Hopkins University Evidence-based Practice Center
NICU	Neonatal Intensive Care Unit
NIH	National Institutes of Health
OGTT	Oral Glucose Tolerance Test
PICO	Population, Intervention, Comparison, and Outcomes
PPG	Postpartum Glucose
RCT	Randomized Controlled Trial
TZD	Thiazolidinedione
WHO	World Health Organization

Appendix A. Results from Step 1 (Identification and Abstraction of Research Gaps from Evidence Report (2008) and Five Published Manuscripts) and Step 2 (Feedback from Authors of 2008 Evidence Report)

Key Question No.	Key Question Topic	Gap Areas	Specific Clarifications/ Additional Gaps Received From Authors of 2008 Evidence Report
I	Benefits and harms of oral diabetes agents as compared to all types of insulin	POPULATIONS • Racial distribution of populations studied needs to be reported. • Glucose tolerance levels at baseline of populations studied needs to be reported. INTERVENTIONS & COMPARISONS • Any 2nd generation sulfonylurea vs. any insulin • Glyburide vs. any insulin • Metformin vs. any insulin • Short-acting insulin vs. long-acting insulin • Short-acting insulin vs. diet alone OUTCOMES • Maternal—Maternal weight, cesarean delivery, preeclampsia, hypoglycemia, perineal tears, operative vaginal delivery, and postpartum hemorrhage • Neonatal—Birth trauma, birth weight, hypoglycemia, shoulder dystocia, congenital malformations, hyperbilirubinemia, large for gestational age, small for gestational age, macrosomia, neonatal intensive care unit admission, and perinatal mortality. SETTINGS • Determining optimal glucose thresholds for medication use in outpatient settings STUDY DESIGNS • Studies with power analyses and larger sample sizes • RCTs with intention-to-treat analyses • Studies with consistent and validated outcome measures across studies • Prospective observational studies with low loss to follow up, adjusting for relevant covariates.	• While many outcome measures were included in these studies, future research needs should include consistent and validated outcome measures that are also clinically relevant. For example, hypoglycemia was defined in several different ways for the different studies, thus limiting the ability to combine results between studies. Hypoglycemia is an objective outcome measure that could be compared between studies. Other such objective, clinically relevant outcome measures should be established for future studies. • For observational studies, practice patterns have evolved over the past few years, with more providers using oral antidiabetic agents. Observational studies may be possible now, but it is critical that adjustment for relevant covariates is included in such studies.

Appendix A. Results from step 1 (identification and abstraction of research gaps from evidence report (2008) and five published manuscripts) and step 2 (feedback from authors of 2008 evidence report) (continued)

Key Question No.	Key Question Topic	Gap Areas	Specific Clarifications/ Additional Gaps Received From Authors of 2008 Evidence Report
II	Benefits and harms of elective cesarean delivery or the choice of timing of induction	POPULATIONS • Women with insulin-requiring (class A2) gestational diabetes INTERVENTIONS & COMPARISONS • Elective labor induction at term vs. expectant management at term • Elective cesarean delivery at term vs. expectant management at term • Elective labor induction or elective cesarean delivery at term vs. expectant management at term OUTCOMES • Maternal – Cesarean delivery, perineal tears, operative vaginal delivery, postpartum hemorrhage, and infection • Neonatal – Shoulder dystocia, congenital malformations, birth trauma, hypoglycemia, hyperbilirubinemia, hypoglycemia, respiratory distress syndrome, neonatal mortality, birth weight, macrosomia, and large for gestational age SETTINGS • At term • Less than 40 weeks gestation SUGGESTED STUDY DESIGNS • Well-designed observational studies with adjustment for other factors influencing labor management (e.g., socio-economic status) with low loss to follow up, adjusting for important covariates • RCTs with intention-to-treat analyses • Stratified analyses if diet-controlled and insulin-controlled patients are included.	• RCTs would be ideal. Only one RCT, with 200 patients has been completed on this topic. Although an RCT might be difficult to perform, it would be an important addition to the literature on this topic and intention-to-treat analyses can be used. • There is also a great need for new prospective observational studies. Most of the previous studies are older and practice patterns have changed over time, making them somewhat less relevant. One important recommendation for future observational studies would be stratification by treatment, such as treatment with diet, insulin, and oral antidiabetic medications. Studies with large numbers of patients, for example at large tertiary care centers, would allow such stratification as well as more consistent treatment within groups. Carrying out the study at a large medical center would also allow the study to be completed in a shorter period of time, thus mitigating the impact of changing practice patterns. • Adjustment for potential confounders is critical in any observational study performed. Because future research in this area will likely depend on observational studies, consistent outcome measures and consideration of confounders will be important to draw conclusions from such studies.

Appendix A. Results from step 1 (identification and abstraction of research gaps from evidence report (2008) and five published manuscripts) and step 2 (feedback from authors of 2008 evidence report) (continued)

Key Question No.	Key Question Topic	Gap Areas	Specific Clarifications/ Additional Gaps Received From Authors of 2008 Evidence Report
III	Risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with gestational diabetes	POPULATIONS • Racial distribution of populations studied need to be consistently reported RISK FACTORS • Maternal lifestyle risk factors• Maternal anthropometry (BMI, weight)• Reproductive factors (e.g., parity) • Contraceptive use (especially progestin-only) • Behavioral factors, including breastfeeding, physical activity, postpartum weight retention • Metabolic risk factors, including insulin sensitivity, HPA stress axis OUTCOMES • Development of type 2 diabetes mellitus• Development of glucose intolerance SETTINGS • Clinical and non-clinical settings SUGGESTED STUDY DESIGNS • Well-designed observational (longitudinal cohort) studies with recruitment at the time of GDM diagnosis and with low loss to follow up. Multivariate analyses adjusting for important covariates needed. • Sampling – random or purposeful sampling is recommended • Consistent definitions for risk factors across studies (e.g. BMI thresholds) • Consistent anthropometric measures across studies • Consistent and accepted definitions for short-term and long-term followup.	• Other risk factors that have not been examined include psychosocial factors (depression, perceived stress, etc.) and comorbidities related to increased risk of type 2 diabetes. • Maternal lifestyle risk factor examples include physical activity and daily caloric intake. • Metabolic risk factors should include subclinical hypercortisolism. • Another risk factor to incude is postpartum depression or other depressive postpartum mental health disorders. • In the outcome section “glucose intolerance” should be “glucose intolerance/prediabetes” • Under suggested study designs the first bulleted point should be updated to the following: “Well-designed observational (longitudinal cohort) studies with recruitment at the time of GDM diagnosis and with low loss to follow up and collection of data on relevant risk factors noted above at the baseline exam on inception of the cohort. • Settings outside of clinical setting should be included since 6 week followups for postpartum women are poor.

Appendix A. Results from step 1 (identification and abstraction of research gaps from evidence report (2008) and five published manuscripts) and step 2 (feedback from authors of 2008 evidence report) (continued)

Key Question No.	Key Question Topic	Gap Areas	Specific Clarifications/ Additional Gaps Received From Authors of 2008 Evidence Report
IV	Performance characteristics (sensitivity, specificity, and reproducibility) of tests for diagnosing type 2 diabetes following a pregnancy with gestational diabetes	POPULATIONS • Racial distribution of populations studied need to be consistently reported INTERVENTIONS & COMPARISONS • Single FBG using ≥7.0 mmol/L threshold vs. OGTT (WHO 1999 criteria) • Other methods for diagnosis including HbA1c vs. OGTT vs. FBG • Home blood glucose monitoring (FBG and random glucose readings) OUTCOMES • Diagnosis of type 2 diabetes mellitus • Reproducibility of tests (single FBG, OGTT using WHO 1999 criteria) SETTINGS • Postpartum period and beyond • What are the optimal frequency screening intervals? (schedule X vs. schedule Y) • Should the screening test and interval of followup vary based on maternal risk factors other than GDM? SUGGESTED STUDY DESIGNS • Longitudinal studies • RCTs of different screening tests (FBG vs. complete OGTT) to determine diagnostic effectiveness of screening test, patient adherence, and patient satisfaction	• Would consider combining suggested study designs to state longitudinal studies and RCTS in one sentence. Could consider adding cluster randomized trials since these are more effectiveness-based than standard RCTs.

Appendix A. Results from step 1 (identification and abstraction of research gaps from evidence report (2008) and five published manuscripts) and step 2 (feedback from authors of 2008 evidence report) (continued)

Key Question No.	Key Question Topic	Gap Areas	Specific Clarifications/ Additional Gaps Received From Authors of 2008 Evidence Report
General Research Gaps (Not Specific to Any Key Question)	-	INTERVENTIONS & COMPARISONS • Effective diagnostic tests for glucose intolerance (not just type 2 diabetes mellitus) in women with history of GDM, even in non-hospitalized patients • Treatment choice (insulin, glyburide, metformin, combinations) and development of type 2 diabetes mellitus • Interventions to improve compliance with postpartum screening for type 2 diabetes mellitus in women with history of GDM. • Studies to develop strategies and technology to track and monitor provider compliance with postpartum glucose tolerance testing in women with history of GDM • RCTs to assess the efficacy of lifestyle interventions in women with history of GDM for prevention of type 2 diabetes mellitus and obesity • RCTs of comparative effectiveness of strategies (medication, behavioral) to prevent obesity, or glucose intolerance (fasting glucose intolerance, insulin resistance, type 2 diabetes mellitus) in women with GDM • Studies to develop strategies and technology to disseminate educational materials and strategies for prevention of type 2 diabetes mellitus in women with history of GDM SETTINGS • Postpartum period • Non-hospitalized patients SUGGESTED STUDY DESIGNS • Observational studies • RCTs.	• Could add after treatment choice thiazolidinediones, exenetide, and januvia to the list.

Abbreviations: BMI=body mass index, GDM=gestational diabetes mellitus, HbA1c =hemoglobin A1c, HPA=hypothalamo-pituitary-adrenal, OGTT=oral glucose tolerance test, RCT=randomized controlled trial, WHO=World Health Organization

Appendix B. Results from Step 3 (Translation of Research Gaps into Researchable Questions) and Step 4 (Online and In-Person Feedback on Research Questions from Local Stakeholders)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
I-1	What is the comparative effectiveness and safety of any second generation sulfonylurea compared to any insulin in the treatment of gestational diabetes with regard to the following maternal and neonatal outcomes? Maternal Outcomes: weight, glycemic control, hypoglycemia, preeclampsia, operative vaginal delivery, hemorrhage, perineal tears, and cesarean delivery Neonatal Outcomes: birth trauma, anoxia, respiratory distress syndrome, birth weight, large for gestational age, macrosomia, small for gestational age, congenital malformations, hyperbilirubinemia, hypoglycemia, NICU admissions, and perinatal mortality	• It is unclear if it intends to evaluate the sulfonlyureas as a class or any specific sulfonylurea. • Rewording suggested to include the following: "How does the effectiveness and safety profile of a ... compare to(with) that of ... in the treatment of ...?"	7.8 (7-9)	6.3 (3-9)	-

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
I-2	What is the comparative effectiveness and safety of metformin compared to any insulin in the treatment of gestational diabetes with regard to the following maternal and neonatal outcomes? Maternal Outcomes: weight, glycemic control, preeclampsia, hemorrhage, operative vaginal delivery, perineal tears, and cesarean delivery Neonatal Outcomes: birth trauma, anoxia, respiratory distress syndrome, birth weight, large for gestational age, macrosomia, small for gestational age, congenital malformations, hyperbilirubinemia, hypoglycemia, NICU admissions, and perinatal mortality	• Rewording suggested to include the following: "How does the effectiveness and safety profile of .... compare to (with) that of .... in the treatment of …?"	7.8 (7-9)	6.7 (5-8)	-
I-3	What is the comparative effectiveness and safety of any oral hypoglycemic medication (i.e., a second generation sulfonylurea or metformin) compared with any insulin in the treatment of gestational diabetes with regard to the following maternal and neonatal outcomes? Maternal Outcomes: cesarean delivery Neonatal Outcomes: birth trauma, fetal weight, and macrosomia	• Rewording suggested to include the following: "How does the effectiveness and safety profile of… compare to (with) that of ... in the treatment of ...?" • Unclear as to why only looking at these limited outcomes in this case.	7.3 (6-8)	7.8 (7-9)	-

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
I-4	What is the comparative effectiveness and safety of a short-acting vs. an intermediate or long-acting insulin in the treatment of gestational diabetes with regard to maternal and neonatal outcomes?	• It is unclear what these outcomes are. • Could you also include rapid-acting insulin (along with short-acting)? • Not clear if you really mean to compare these to each other or within each type of insulin. • Previous questions delineated the maternal and neonatal outcomes of interest. It would be helpful to see specifically which of the outcomes may be impacted by these research questions. • Need to clarify the outcomes.	6.3 (3-8)	5.5 (3-8)	• This question is not clinically relevant because short and intermediate insulin may be combined. Also regular insulin vs. insulin lispro would be a good comparison, but it is not related to key question I.
I-5	What is the comparative effectiveness and safety of a short-acting insulin compared to diet alone in the treatment of gestational diabetes with regard to maternal and neonatal outcomes?	• Need to define the outcome of interest. • Rewording suggested to include: "How does the effectiveness and safety of ... compare to (with) that of ...?" • Need to clarify outcomes.	6.7 (5-8)	5 (3-8)	• If someone chooses diet alone they will not choose short-acting insulin.

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
I-6	What is the comparative effectiveness and safety of oral hypoglycemic agents like thiazolidinediones (TZDs), exenetide, and sitagliptin in the treatment of gestational diabetes with regard to maternal and neonatal outcomes?	• Specify the nature of outcomes—short-term vs. long-term, surrogate outcomes vs. clinical outcomes vs. patient oriented outcomes. • Not all these are considered "oral hypoglycemic agents." Exenatide is an injectable non-insulin agent. • Need to clarify outcomes.	7.7 (7-9)	5.2 (1-8)	• There are serious feasibility issues with using newer agents. Glyburide has been shown to not cross the placenta, thus it is acceptable to use in an RCT. • In order to begin using newer agents there needs to be a build-up of their evidence base.

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
I-Outcomes of interest	Maternal outcomes cesarean delivery, glycemic control (FBG), 1-hr and 2-hr postprandial glucose (PPG), hemorrhage, hypoglycemia, operative vaginal delivery, perineal tears, preeclampsia, weight Neonatal outcomes anoxia, birth trauma, birth weight, congenital malformations, hyperbilirubinemia, hypoglycemia, large for gestational age, macrosomia, mortality, neonatal intensive care admissions, respiratory distress syndrome, shoulder dystocia, small for gestational age. (Note—These outcomes were not included in the online feedback process. They were discussed in the in-person meeting).	N/A	N/A	N/A	• Patient centered outcomes needed. • Long-term offspring outcomes/ developmental outcomes are needed. • Shoulder dystocia is a maternal outcome. • Shoulder dystocia is subjective. • The outcomes collected across studies vary. • Some outcomes should be split out. • Operative vaginal delivery is subjective. • Cesarean delivery should be changed to “cesarean delivery indication”. • Should look at how many vaginal attempts resulted in cesarean deliveries and why cesarean deliveries were the chosen option (complications, etc.). • In regards to glycemic control, should look at the difficulty in maintaining glycemic control (dosing). • In regards to birth trauma, should also take into consideration bone fatigue specifications.

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
I-Study needs and challenges	Populations • Racial distribution of study participants Study Designs • RCTs Analyses • Intention-to-treat analyses• Race-specific analyses Other Issues • Optimal glucose thresholds for initiating therapy? • Consistent reporting of baseline glucose tolerance levels Outcomes • Consistent ascertainment, definition, and measurement of outcomes (across studies) • Short-Term—Neonatal and maternal hypoglycemia, birth trauma, perineal tears, postpartum hemorrhage, etc. • Long-Term—Infantile growth trajectory, neonatal & maternal metabolic alterations & adiposity, etc. (Note—These study needs and challenges were not included in the online feedback process. They were discussed in the in-person meeting).	N/A (Note—These study needs and challenges were not included in the online feedback process. They were discussed in the in-person meeting).	N/A	N/A	• Additional suggestions included stratification by socio-economic status, education level, access to care, and screening. • Explain trajectory of patients and possibly classify the severity of GDM through C-peptide levels, HbA1c, etc.

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
II-1	What is the comparative effectiveness and safety of elective labor induction vs. expectant management at term in the management of labor in women with gestational diabetes with regard to the following maternal and neonatal outcomes? Maternal Outcomes: cesarean delivery, operative vaginal delivery, hemorrhage, perineal tears, and infection Neonatal Outcomes: birth trauma, anoxia, congenital malformations, hyperbilirubinemia, hypoglycemia, respiratory distress syndrome, shoulder dystocia, neonatal mortality, birth weight, large for gestational age, macrosomia, small for gestational age, and NICU admissions	• Rewording suggested to include the following: "In planning and managing delivery of women with gestational diabetes at term, how does the effectiveness and safety of elective labor induction compare to that of expectant management with regard to ...?"	7.7 (7-8)	5.3 (3-7)	• Should use 'delivery' vs. 'labor' in the wording of the question. • Elective assumes that all other factors are OK, thus the following change should be made: “elective (with no complications)”. • A previous cesarean delivery would exclude you from the study with the current practices.• The term “at term” is vague. Should define by gestational age, term equals 40 weeks.

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
II-2	What is the comparative effectiveness and safety of elective cesarean delivery vs. expectant management at term in the management of labor in women with gestational diabetes with regard to the following maternal and neonatal outcomes? Maternal Outcomes: hemorrhage and infection Neonatal Outcomes: birth trauma, anoxia, congenital malformations, hyperbilirubinemia, hypoglycemia, respiratory distress syndrome, shoulder dystocia, neonatal mortality, birth weight, large for gestational age, macrosomia, small for gestational age, and NICU admissions	• Rewording suggested to replace "management of labor" with "planning and management of delivery at term." • Suggest adding wound complications for maternal outcomes. Any interest in longer term outcomes (e.g., maternal incontinence, previa/acreta in future pregnancies, longer term baby outcomes like obesity, diabetes, cognitive function)?	7.2 (6-8)	5.8 (4-7)	• Stratify by EFW and by maternal weight.
II-3	What is the comparative effectiveness and safety of elective labor induction or cesarean delivery compared to expectant management at term in the management of labor in women with insulin-requiring (class A2) gestational diabetes with regard to the following maternal and neonatal outcomes? Maternal Outcome: non-elective cesarean delivery (if elective induction is performed) Neonatal Outcome: macrosomia	• Using "comparative effectiveness" and "compared to" in same sentence is redundant and awkward; also suggest rewording to replace "management of labor" with "planning and management of delivery at term." • May want to look at more outcomes (especially neonatal).	7.2 (6-8)	6 (4-7)	N/A (This question was not discussed in the in-person meeting.)

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
II-Outcomes of interest (for II-1 & II-2 only)	Maternal outcomes cesarean delivery, operative vaginal delivery, hemorrhage, perineal tears, infection, non-elective cesarean delivery (if elective induction is performed) Neonatal outcomes anoxia, birth trauma, birth weight, congenital malformations, hyperbilirubinemia, hypoglycemia, large for gestational age, macrosomia, mortality, NICU admissions, respiratory distress syndrome, shoulder dystocia, small for gestational age (Note—These outcomes were not included in the online feedback process. They were discussed in the in-person meeting).	N/A	N/A	N/A	• Patient-centered outcomes needed. • Long-term offspring outcomes / developmental outcomes are needed. • Shoulder dystocia is a maternal outcome. • Long-term studies collecting neonatal developmental outcomes are feasible. • Shoulder dystocia is subjective. • Some outcomes should be split out. • Operative vaginal delivery is subjective. • Cesarean delivery under the maternal outcomes heading should be changed to "indication for cesarean delivery." • Should look at how many vaginal attempts resulted in cesarean deliveries and why cesarean deliveries were the chosen option (complications, etc.). • In regards to glycemic control, should look at the difficulty in maintaining glycemic control (dosing). • In regards to birth trauma, should also take into consideration bone fatigue specifications.

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
II- Study needs and challenges (for II-1 & II-2 only)	Study Designs • Observational studies with adjustment for factors that affect labor management (e.g., socio-economic status) • RCTs/observational studies to address cesarean delivery based on EFW or gestational age Analyses • Stratified analyses (if diet-controlled and insulin-controlled women with GDM are included) (Note—These study needs and challenges were not included in the online feedback process. They were discussed in the in-person meeting).	N/A (Note—These study needs and challenges were not included in the online feedback process. They were discussed in the in-person meeting).	N/A	N/A	• Will have problems with providers not following the random allocation. • In larger populations crossing between arms is negligible. • Can do cluster randomization.

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
III-1	What is the evidence that presence of maternal lifestyle risk factors (e.g., behavioral risk factors, breastfeeding, physical activity, daily caloric intake) are associated with short-term and long-term development of type 2 diabetes mellitus or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	• Maybe focus on factors that accelerate progression to Type 2 diabetes mellitus, as women with GDM are already at risk. • Suggest rewording to: "How strongly do specific elements of maternal lifestyle (...) correlate with short- and long-term development ...?" or "Following a pregnancy complicated by GDM, what is the impact of specific elements of maternal lifestyle (e.g., ...) on both short- and long-term development of ..?"	8.2 (7-9)	6.2 (4-7)	During the in-person meeting all research questions related to key question III were presented together. Thus comments were made on the questions as a whole, with specific questions being addressed when necessary. The comments were as follows: • In addition to psychosocial factors should also look at anxiety disorders and eating disorders. • Should also look at interpregnancy interval. • Should look at the number of GDM-associated pregnancies of each person. • Family history may be difficult to collect from individuals who do not know their family history. • Should look at racial/ethnic groups.

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
III-2	What is the evidence that presence of maternal psychosocial risk factors (e.g. depression, postpartum depression, other depressive postpartum mental health disorders, perceived stress, etc.) are associated with short-term and long-term development of type 2 diabetes mellitus or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	• Maybe add "accelerated progression" to short-term and long-term type 2 diabetes. • See suggested wording for previous question... except that unlike in the other question which include examples of both positive and negative (or neutral) lifestyle elements and therefore are incorrectly being termed "risk factors," here the examples are all risk factors. You could phrase the question as: "To what extent does the presence of maternal psychosocial risk factors increase the risk of developing ...?"	7.5 (6-8)	5.2 (5-6)	-

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
III-3	What is the evidence that presence of maternal metabolic risk factors (e.g. insulin sensitivity, HPA axis stress (subclinical hypercortisolism)) are associated with short-term and long-term development of type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	• Maybe specify the timing of these risk factor assessments. • Suggested rewording: "To what extent does the presence of maternal metabolic risk factors (e.g. ,...) increase the short- and long-term risk of developing ...?" • Are the maternal risk factors present before, during, or after pregnancy?	6.8 (5-9)	5.2 (3-7)	-
III-4	What is the evidence that presence of maternal co-morbidities (e.g., obesity, hypertension, hypercholesterolemia) are associated with short-term and long-term development of type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	• Specifiy whether maternal comorbidities are during pregnancy, before, or after.	8 (7-9)	7 (6-9)	-

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
III-5	What is the evidence that progestin-only contraceptives are associated with short-term and long-term development of type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	• Compared to nonuse. • Suggested rewording: "Does use of progestin-only contraceptives following a pregnancy complicated by GDM increase the risk of short- and long-term development of ...." • Are there parameters for short- and long-term development of type 2 diabetes?	7.7 (7-9)	6.7 (5-8)	-

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
III-Study needs and challenges	Populations • Consistency in race/ethnicity (across studies) • Need for recruitment at time of GDM diagnosis Study Designs • Observational studies with ascertainment of covariates and confounders • Random or purposeful sampling for more representative samples Analyses • Multivariate analyses for pooled estimates of risk Other Issues • Consistency in risk factors and their definitions (across studies) • Focus on specific categories of anthropometry (e.g., BMI, weight) and reproductive factors (e.g., parity) (Note—These study needs and challenges were not included in the online feedback process. They were discussed in the in-person meeting).	N/A (Note—These study needs and challenges were not included in the online feedback process. They were discussed in the in-person meeting).	N/A	N/A	• Current evidence is of poor quality. • Studies were short-term and had low adjustment levels.

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
IV-1	What is the accuracy of a single fasting blood glucose test compared to the full 2-hour 75-gm OGTT in diagnosing type 2 diabetes following a pregnancy with gestational diabetes? Does the accuracy of the fasting blood glucose test compared to the full 2-hour 75-gm OGTT vary with the postpartum testing interval in diagnosing type 2 diabetes following a pregnancy with gestational diabetes?	• Suggested rewording: "How does a single fasting .... compare to the full ... in detecting ...? How does the postpartum testing interval affect the comparative detection rates between a single fasting ... and the full ...?"	7.7 (6-9)	6.8 (5-8)	• Compliance is an issue. If a patient will only use the less accurate test than that is better to collect data from the less accurate test vs. obtain no test results.
IV-2	What is the accuracy of the HbA1c test compared to the 2-hour 75-gm OGTT or single fasting blood glucose test in diagnosing type 2 diabetes following a pregnancy with gestational diabetes?	• See previous wording suggestions	7 (5-9)	6.2 (4-8)	• No studies looked at HbA1c.
IV-3	What is the reproducibility of the 2-hour 75-gram OGTT vs. a single fasting blood glucose test vs. a single HbA1c test in diagnosing type 2 diabetes following a pregnancy with gestational diabetes?	• Unclear how the three tests are being compared and which is the gold standard? • Unclear what you mean by reproducibility ... do you mean consistency of results if repeated within the same patient? If so, then the last part of the sentence referring to "diagnosis" is unnecessary.	7.2 (5-9)	6.2 (3-8)	Not discussed at the in-person meeting. Only included in the online questionnaire.

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
IV-Study needs and challenges (for IV-1 & IV-2 only)	Populations • Consistency in terms of race/ethnicity and future diabetes risk (family history of type 2 diabetes, prior GDM, etc.) (across studies) • Consistent protocols for recruitment of participants (across studies) Other Issues • Do the screening test and interval of followup vary based on maternal risk factors other than GDM? (Note—These study needs and challenges were not included in the online feedback process. They were discussed in the in-person meeting).	N/A (Note—These study needs and challenges were not included in the online feedback process. They were discussed in the in-person meeting).	N/A	N/A	-
Additional Overall Research Questions	-	• Should women with a history of GDM be treated as "once a GDM always a GDM" or should they be re-screened at each new pregnancy? What innovations in "delivery of care" models (e.g., "group care" within the clinical setting, alterations of frequency in visits) might improve outcomes of GDM? How does the relative "volatility" of glycemic control (even with same HbA1c) affect outcomes?	-	-	• Should include patient-centered outcomes like patient preference. • Compliance is also important. • Diet should also be taken into consideration including access to whole foods vs. packaged foods. • Outcomes may be improved using group prenatal care which is used in the care of pregnant women without GDM. In support groups, the women follow each other's progress. • May improve outcomes by using an alternative clinic space and time. You could have a group of women meet at a gym and do exercise while they are waiting for their appointment.

Appendix B. Results from step 3 (translation of research gaps into researchable questions) and step 4 (online and in-person feedback on research questions from local stakeholders) (continued)

Research Question Number	Initial Research Questions (As provided to local stakeholders for Online Feedback, Unless otherwise specified)*	Comments from Online Feedback	Online Feedback—Likely Clinical Benefit/ Importance Mean (Range) on a scale of 1-9	Online Feedback—Likely Ability to Conduct a Study Mean (Range) on a scale of 1-9	Comments from In-Person Meeting
		• If interested in progestin-only pills in women after GDM, it might be worth looking at other contraceptive methods (e.g., implants, intrauterine devices) • In addition to expanding psychosocial risk factors as listed earlier, access to care and health care disparities are other areas perhaps offering a relationship between GDM and the development of type 2 diabetes.			• Must take into consideration transportation and healthcare disparities. • Postpartum depression may be helped by followup visits because patients who require attention get it. • No one has used information technology tools to track compliance.

* The wording of research questions and outcomes in this table and footnotes reflects the wording of research questions and outcomes as provided to the local stakeholders in the online feedback process. The wording of some research questions and outcomes was modified as appropriate before and after the in-person meeting on the basis of feedback received from the local stakeholders. The next version of wording of research questions (Delphi round 1) is provided in Appendix C.

Abbreviations: BMI=body mass index, EFW=estimated fetal weight, FBG=fasting blood glucose, GDM=gestational diabetes mellitus, HbA1c=hemoglobin A1c, HPA=hypothalamo-pituitary-adrenal, NICU=neonatal intensive care unit, OGTT=oral glucose tolerance test, PPG=postpartum glucose, RCT=randomized controlled trial, TZD=thiazoledinedione

Appendix C. Results from Step 5 (Online Feedback, Consensus Development, and Prioritization of Research Questions by External Stakeholders (Delphi Approach))

Key Question Number	Key Question Topic	Research Question Number	Research Question (as entered into Delphi Round 1)*	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 1	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 2	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 3	Included in Final List of Research Questions of High Clinical Benefit/ Importance?
I	Benefits and harms of oral diabetes agents as compared to all types of insulin	I-1	What is the effectiveness and safety of any of the second generation sulfonylureas compared to any insulin in the treatment of gestational diabetes with regard to the following maternal, neonatal, and long-term offspring outcomes?**	Mean=8.2 Range=7 to 9 (Consensus - HIGH IMPORTANCE)	N/A****	N/A****	√
I	Benefits and harms of oral diabetes agents as compared to all types of insulin	I-2	What is the effectiveness and safety of metformin compared to any insulin in the treatment of gestational diabetes with regard to the following maternal, neonatal, and long-term offspring outcomes?**	Mean=7.9 Range=6 to 9 (Consensus - HIGH IMPORTANCE)	N/A****	N/A****	√
I	Benefits and harms of oral diabetes agents as compared to all types of insulin	I-3	What is the effectiveness and safety of individual insulin regimens (e.g., long or intermediate acting) or combination regimens (e.g., long + short-acting) compared to other insulin regimens or combination regimens in the treatment of gestational diabetes with regard to the following maternal, neonatal , and long-term offspring outcomes?**	Mean=6.4 Range=4 to 8 (No Consensus)	Mean=7.3 Range=6 to 9 (Consensus - HIGH IMPORTANCE)	N/A****	√

Appendix C. Results from step 5 (online feedback, consensus development, and prioritization of research questions by external stakeholders (Delphi approach)) (continued)

Key Question Number	Key Question Topic	Research Question Number	Research Question (as entered into Delphi Round 1)*	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 1	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 2	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 3	Included in Final List of Research Questions of High Clinical Benefit/ Importance?
I	Benefits and harms of oral diabetes agents as compared to all types of insulin	I-4	What is the effectiveness and safety of other hypoglycemic drug classes (e.g., thiazolidinediones, DPP-4 inhibitors, GLP-1 agonists, meglitinides) compared to any insulin in the treatment of gestational diabetes with regard to the following maternal, neonatal, and long-term offspring outcomes?**	Mean=5.8 Range=3 to 8 (No Consensus)	Mean=6.7 Range=3 to 9 (No Consensus)	Mean=6.9 Range=4 to 9 (Consensus - HIGH IMPORTANCE)	√
II	Benefits and harms of elective cesarean delivery or the choice of timing of induction	II-1	What is the effectiveness and safety of elective labor induction at 40 weeks compared to expectant management in women with gestational diabetes with regard to the following maternal and neonatal outcomes?*** <u>Populations of Interest:</u> • All women with gestational diabetes • Women with insulin-requiring (class A2) gestational diabetes • Obese women with gestational diabetes • Women with gestational diabetes with high estimated fetal weight (e.g., >4000 or >4500 gram) • Women with different parities • Women of various races/ethnicities	Mean=7.8 Range=6 to 9 (Consensus - HIGH IMPORTANCE)	N/A****	N/A****	√

Appendix C. Results from step 5 (online feedback, consensus development, and prioritization of research questions by external stakeholders (Delphi approach)) (continued)

Key Question Number	Key Question Topic	Research Question Number	Research Question (as entered into Delphi Round 1)*	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 1	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 2	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 3	Included in Final List of Research Questions of High Clinical Benefit/ Importance?
II	Benefits and harms of elective cesarean delivery or the choice of timing of induction	II-2	What is the effectiveness and safety of elective cesarean delivery at 40 weeks compared to expectant management in women with gestational diabetes* with regard to the following maternal and neonatal outcomes?*** Populations of Interest: • All women with gestational diabetes • Women with insulin-requiring (class A2) gestational diabetes • Obese women with gestational diabetes • Women with gestational diabetes with high estimated fetal weight (e.g., >4000 or >4500 gram) • Women with different parities • Women of various races/ethnicities	Mean=7.3 Range=4 to 9 (Consensus - HIGH IMPORTANCE)	N/A****	N/A****	√
III	Risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with gestational diabetes	III-1	What is the evidence that maternal health behaviors (e.g., breastfeeding, physical activity, diet) are associated with the risk of developing type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	Mean=8.1Range=7 to 9(Consensus - HIGH IMPORTANCE)	N/A****	N/A****	√

Appendix C. Results from step 5 (online feedback, consensus development, and prioritization of research questions by external stakeholders (Delphi approach)) (continued)

Key Question Number	Key Question Topic	Research Question Number	Research Question (as entered into Delphi Round 1)*	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 1	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 2	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 3	Included in Final List of Research Questions of High Clinical Benefit/ Importance?
III	Risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with gestational diabetes	III-2	What is the evidence that maternal psychosocial factors (e.g., mood disorders, eating disorders, stress) are associated with the risk of developing type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	Mean=6.0 Range=4 to 9 (No Consensus)	Mean=6.0 Range=4 to 8 (No Consensus)	Mean=6.0 Range=5 to 8 (No Consensus)	
III	Risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with gestational diabetes	III-3	What is the evidence that maternal metabolic measures (e.g., fasting insulin levels, OGTT measures, HPA axis stress (subclinical hypercortisolism)) are associated with the risk of developing type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	Mean=6.6 Range=3 to 9 (No Consensus)	Mean=7.3 Range=6 to 8 (Consensus - HIGH IMPORTANCE)	N/A****	√
III	Risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with gestational diabetes	III-4	What is the evidence that pre-pregnancy, antenatal, and postpartum co-morbid conditions (e.g., obesity, hypertension, hypercholesterolemia) are associated with the risk of developing type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	Mean=7.3 Range=5 to 9 (No Consensus)	Mean=7.4 Range=6 to 9 (Consensus - HIGH IMPORTANCE)	N/A****	√

Appendix C. Results from step 5 (online feedback, consensus development, and prioritization of research questions by external stakeholders (Delphi approach)) (continued)

Key Question Number	Key Question Topic	Research Question Number	Research Question (as entered into Delphi Round 1)*	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 1	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 2	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 3	Included in Final List of Research Questions of High Clinical Benefit/ Importance?
III	Risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with gestational diabetes	III-5	What is the evidence that contraceptive method (e.g., depo-provera) is associated with the risk of developing type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	Mean=4.9 Range=2 to 9 (No Consensus)	Mean=5.7 Range=3 to 8 (No Consensus)	Mean=5.9 Range=4 to 8 (No Consensus)	
III	Risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with gestational diabetes	III-6	What is the evidence that the inter-conception interval is associated with the risk of developing type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	Mean=5.6Range=4 to 7(No Consensus)	Mean=5.3Range=2 to 8(Consensus - MEDIUM IMPORTANCE)	N/A****	
III	Risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with gestational diabetes	III-7	What is the evidence that family history, gene mutations, genotypes, gene-environment interactions, epigenetic modifications, or other biomarkers are associated with the risk of developing type 2 diabetes among women with gestational diabetes? Are there differences in these associations by race or ethnic group?	Mean=7.4 Range=3 to 9 (Consensus - HIGH IMPORTANCE)	N/A****	N/A****	√

Appendix C. Results from step 5 (online feedback, consensus development, and prioritization of research questions by external stakeholders (Delphi approach)) (continued)

Key Question Number	Key Question Topic	Research Question Number	Research Question (as entered into Delphi Round 1)*	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 1	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 2	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 3	Included in Final List of Research Questions of High Clinical Benefit/ Importance?
III	Risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with gestational diabetes	III-8	What is the comparative effectiveness of various lifestyle interventions for prevention of type 2 diabetes mellitus, glucose intolerance, and obesity in women with a history of gestational diabetes?	Mean=7.7 Range=6 to 9 (Consensus - HIGH IMPORTANCE)	N/A****	N/A****	√
III	Risk factors associated with short-term and long-term development of type 2 diabetes following a pregnancy with gestational diabetes	III-9	What is the comparative effectiveness of various innovative strategies and technologies to disseminate educational materials on prevention of type 2 diabetes mellitus and glucose intolerance to women with a history of gestational diabetes?	Mean=6.6 Range=1 to 9 (No Consensus)	Mean=7.3 Range=2 to 9 (Consensus - HIGH IMPORTANCE)	N/A****	√

Appendix C. Results from step 5 (online feedback, consensus development, and prioritization of research questions by external stakeholders (Delphi approach)) (continued)

Key Question Number	Key Question Topic	Research Question Number	Research Question (as entered into Delphi Round 1)*	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 1	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 2	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 3	Included in Final List of Research Questions of High Clinical Benefit/ Importance?
IV	Performance characteristics (sensitivity, specificity, and reproducibility) of tests for diagnosing type 2 diabetes following a pregnancy with gestational diabetes	IV-1	What are the performance characteristics (sensitivity, specificity, and reproducibility) of a single fasting blood glucose test compared to the full 2-hour 75-gm OGTT in screening for type 2 diabetes following a pregnancy with gestational diabetes? Does the accuracy of the fasting blood glucose test compared to the full 2-hour 75-gm OGTT vary with the postpartum testing interval in screening for type 2 diabetes following a pregnancy with gestational diabetes?	Mean=5.9 Range=1 to 9 (No Consensus)	Mean=6.6 Range=4 to 9 (No Consensus)	Mean=6.7 Range=1 to 9 (Consensus - HIGH IMPORTANCE)	√
IV	Performance characteristics (sensitivity, specificity, and reproducibility) of tests for diagnosing type 2 diabetes following a pregnancy with gestational diabetes	IV-2	What are the performance characteristics (sensitivity, specificity, and reproducibility) of the HbA1c test compared to the 2-hour 75-gm OGTT in screening for type 2 diabetes following a pregnancy with gestational diabetes?	Mean=7.3 Range=5 to 9 (No Consensus)	Mean=7.9 Range=6 to 9 (Consensus - HIGH IMPORTANCE)	N/A****	√

Appendix C. Results from step 5 (online feedback, consensus development, and prioritization of research questions by external stakeholders (Delphi approach)) (continued)

Key Question Number	Key Question Topic	Research Question Number	Research Question (as entered into Delphi Round 1)*	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 1	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 2	Clinical Benefit/ Importance (Mean, Range on a Scale of 1-9) Delphi Round 3	Included in Final List of Research Questions of High Clinical Benefit/ Importance?
IV	Performance characteristics (sensitivity, specificity, and reproducibility) of tests for diagnosing type 2 diabetes following a pregnancy with gestational diabetes	IV-3	What is the comparative effectiveness of various strategies or interventions to improve clinician compliance with postpartum screening guidelines for type 2 diabetes mellitus and glucose intolerance in women with a history of gestational diabetes?	Mean=7.8 Range=5 to 9 (Consensus - HIGH IMPORTANCE)	N/A****	N/A****	√
IV	Performance characteristics (sensitivity, specificity, and reproducibility) of tests for diagnosing type 2 diabetes following a pregnancy with gestational diabetes	IV-4	What is the comparative effectiveness of various innovative strategies or technologies (e.g., electronic health records) to track postpartum screening for the development of type 2 diabetes mellitus and glucose intolerance in women with a history of gestational diabetes?	Mean=6.6 Range=4 to 9 (No Consensus)	Mean=6.2 Range=2 to 9 (No Consensus)	Mean=5.8 Range=3 to 8 (No Consensus)	

* The wording of research questions and outcomes in this table and footnotes reflects the wording of research questions and outcomes as entered into Delphi round 1. The wording of some research questions and outcomes was modified as appropriate during Delphi rounds 1, 2, and/or 3 on the basis of feedback from the external stakeholders. The final wording of highly important research questions is provided in Table 6. The final wording of research questions that did not make the final list of highly important research questions is provided in Table 3.

Appendix C. Results from step 5 (online feedback, consensus development, and prioritization of research questions by external stakeholders (Delphi approach)) (continued)

** Outcomes for Research Questions I-1, I-2, I-3, and I-4

Maternal Outcomes: mortality, cesarean delivery (including primary cesarean and repeat cesarean), indication for cesarean delivery, complications of cesarean delivery (e.g., wound infection, wound dehiscence), vaginal delivery (and specify type: spontaneous or operative), perineal tears, hemorrhage, hypoglycemia, shoulder dystocia, pregnancy weight gain, glycemic control (e.g., fasting blood glucose, 2-hr postprandial glucose, development of postpartum type 2 diabetes mellitus, hypertensive disorders of pregnancy (e.g., gestational hypertension, preeclampsia), patient-reported outcomes (e.g., patient treatment preference, quality of life), and medication adherence

Neonatal Outcomes: mortality, anoxia, birth trauma (e.g., bone fractures, brachial plexus palsy), birth weight, hyperbilirubinemia, hypoglycemia, large for gestational age and macrosomia, small for gestational age, macrosomia, neonatal intensive care admission, and respiratory distress syndrome

Long-Term Offspring Outcomes: growth, anthropometrics, and chronic diseases (e.g., obesity, diabetes).

*** Outcomes for Research Questions II-1 and II-2

Maternal Outcomes: mortality, cesarean delivery (including primary cesarean and repeat cesarean), indication for cesarean delivery, complications of cesarean delivery (e.g., wound infection, wound dehiscence), vaginal delivery (spontaneous, operative), perineal tears, hemorrhage, patient-reported outcomes (e.g., patient preference , quality of life), adherence, length of hospital stay, and pulmonary embolism

Neonatal Outcomes: mortality, anoxia, birth trauma (e.g., bone fractures, brachial plexus palsy), birth weight, hyperbilirubinemia, hypoglycemia, large for gestational age and macrosomia, small for gestational age, neonatal intensive care admission, and respiratory distress syndrome.

**** N/A=Not applicable. Consensus was already achieved in a previous Delphi round.

Abbreviations: DPP-4=dipeptidyl peptidase-4, GLP-1=glucagon-like peptide-1, HbA1c=hemoglobin A1c, HPA=hypothalamo-pituitary-adrenal, OGTT=oral glucose tolerance test.

Appendix D. Results and Comments from External Stakeholders (Delphi Round 1)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
I-1	What is the effectiveness and safety of any of the second generation sulfonylureas compared to any insulin in the treatment of gestational diabetes with regard to the following maternal, neonatal, and long-term offspring outcomes?	Question A: Is this research question clearly worded?	No. Consider dividing maternal outcomes into short-term and long-term, as you did for neonatal and long-term offspring outcomes.	Yes	No. Restate and reorder maternal outcomes. If medication adherence is not kept, none of the other outcomes are valid. Follow with glycemic control, patient reported outcomes, pregnancy-related issues, then delivery issues, and end with mortality.	Yes	Yes	Yes	Worded clearly. I have some suggestions as to things that are missing. Effectiveness and safety are mentioned, what about efficacy? If safety is being investigated I would expect efficacy to also be included. I expect that prediabetes as well as diabetes are included in the postpartum outcomes for the mother. Should outcomes for neonate include early/late preterm?	No. Neonatal outcomes - need to list macrosomia only once (delete second reference after small for gestational age). Define "macrosomia" specifying birth weight cutoff that refers to term births. Define the interval for "postpartum" type 2 diabetes— <12 months? Add postpartum weight retention to maternal outcomes.	No. I think it should be 'What are', not 'What is...' For neonatal outcomes, specify the period of interest for 'mortality': I would suggest 'perinatal' mortality, to incorporate the outcome of stillbirth. Also, suggest broadening the outcome of 'anoxia,' as lesser degrees of oxygen deprivation are of relevance. Perhaps 'hypoxia/anoxia'.	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	9	9	8	7	8	7	9	8	8.2	7	9

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question C: Rationale for above rating	The long-term outcomes have not been studied as well as short-term outcomes, for obvious reasons (difficult to have adequate retention of cohort).	The practice of utilizing oral agents to treat GDM is widespread. Without robust evidence regarding their impact on the health of the neonates, this widespread use is concerning.	This question seems to be a central basis for the study.	If a second generation sulfonylurea is safe and beneficial for mother and offspring, treatment would be less onerous and compliance is likely to improve. I believe that the biggest question is safety. Since glyburide crosses the placenta according to a recent study which found low levels of glyburide in cord blood (but cord levels were 70% of the low levels in maternal blood measured simultaneously), we need to know whether fetal and long-term effects are good, bad, or neutral. Measurements of glucose disappearance curves in neonates exposed and not exposed in utero may be helpful. The biggest concern is in-utero programming, and current studies really do not address that ... it will take very long-term studies to do so.	The question has been asked before but only in a small number of studies. An RCT comparing oral agents and insulin regimens would be important.	Research is very important since a good outcome for oral agents would provide an alternative for insulin.	-	Perinatal and long-term outcomes for women and their offspring are necessary to develop evidence-based recommendations of treatments during GDM pregnancy.	-	N/A	N/A	N/A
I-2	What is the effectiveness and safety of metformin compared to any insulin in the treatment of gestational diabetes with regard to the following maternal, neonatal, and long-term offspring outcomes?	Question A: Is this research question clearly worded?	Yes, although please see comments for the previous questions.	Yes	No. Reword maternal outcome as in Question 1 with patient adherence first followed by pregnancy considerations and ending with delivery and mortality.	Yes	Yes	No. Macrosomia is listed twice - once with large for gestational age and once alone.	No. Same as 1-1	No. Comments similar to previous page.	No. Identical comments as for Question I-1.	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	9	8	8	6	8	6	9	8	7.9	6	9

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question C: Rationale for above rating	Metformin has been shown to delay cardiovascular disease or macrovascular complications but given outside of pregnancy; treatment during pregnancy would be interesting.	Impact on neonatal outcomes and particularly long-term outcomes in needed.	-	Answer would be similar to that for glyburide in question A, except that metformin crosses the placenta to a much greater extent than glyburide. Being an insulin sensitizer, metformin might also have long-term metabolic effects on the offspring.	See prior comment.	-	-	Metformin safety and efficacy has not been firmly established based on available studies.	-	N/A	N/A	N/A
I-3	What is the effectiveness and safety of individual insulin regimens (e.g. long or intermediate acting) or combination regimens (e.g. long + short-acting) compared to other insulin regimens or combination regimens in the treatment of gestational diabetes with regard to the following maternal, neonatal , and long-term offspring outcomes?	Question A: Is this research question clearly worded?	No. See previous questions. Also, administration route is important and should be included--so, short-acting but also subcutaneous insulin infusions in addition to type of insulin. Additionally, glycemic control might not be measured by the measures you mention—what about fructosamine or glycated albumin? Fasting and 2-hour glucose values fluctuate so much day to day and that is the reason that HgbA1c is recommended for testing postpartum now, although HgbA1c is obviously not ideal in pregnancy due to its long half-life.	Yes	No. The question is unclear as far as what you are comparing. Are you comparing single long or intermediate acting insulin injection given at a regular interval to a long acting insulin supplemented with short acting insulin titrated by blood sugar?	Yes	Yes	Yes	No. Same as 1-1 for issues of efficacy and outcome specification. Not clear what 'other insulin regimens' would include. Research question should be more specific in defining comparison group.	No. Similar to previous questions. Very important to add postpartum weight retention. Reword pregnancy weight gain and use "gestational weight gain". The weight gain before and after treatment should be separately evaluated. Glycemic control may involve 1 hr or 2 hr postprandial data.	No. "What are the safety and effectiveness...". same comments about neonatal outcomes. All the different examples are confusing. Could this be simplified "What are the comparative effectiveness and safety of various individual insulin regimens..."?	N/A	N/A	N/A

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question B: Likely clinical benefit? (scale=1 to 9)	6	4	6	6	8	8	7	5	8	6.4	5	8
		Question C: Rationale for above rating	Administration methods are likely to be studied in near future.	I am not sure "one size" will fit all.	Really unsure of the question, however this could be very important if it is identifying the most effective insulin dosing regimen.	While of great interest, the issues of the best insulin regimen are a little less pressing than the use of oral agents. We do need to know whether and to what extent some of the newer insulin analogs cross the placenta and whether that might have an effect on the offspring.	It would be useful to compare different insulin regimens during pregnancy. Most studies have been in nonpregnant patients.	-	-	Much more data exist regarding the safety and efficacy of insulin treatment for GDM. The specific regimens are adapted for individual glucose control.	-	N/A	N/A	N/A
I-4	What is the effectiveness and safety of other hypoglycemic drug classes (e.g. thiazolidinediones, DPP-4 inhibitors, GLP1 agonists, meglitinides) compared to any insulin in the treatment of gestational diabetes with regard to the following maternal, neonatal, and long-term offspring outcomes?	Question A: Is this research question clearly worded?	No. I am not sure that lumping these drugs into a single category is that useful? Do you want to consider each separately?	Yes	Yes	Yes	Yes	Yes	No. Will any insulin treatment be too heterogeneous to provide any useful information? Given the differences in both groups this comparison will be difficult to sort out.	No. Same comments related to outcomes statements.	No. Same comments as for question I-1.	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	5	3	8	7	7	8	5	4	5	5.8	3	8
		Question C: Rationale for above rating	-	I am not sure we should more forward with other agents until the issue has been resolved with the two most commonly used.	-	If any of these agents are safe and effective in pregnancy it would be terrific. However, currently there is very little information on the most basic of issues - do they cross the placenta at various times in gestation? Before trials in human are undertaken I believe animal models should be investigated. The action of many of these drugs is very complex and fetal effects would be quite unpredictable without animal models first.	-	-	-	These questions are beneficial only if other oral medications that have been widely used in humans for diabetes management and have limited transport across the placenta do not demonstrate efficacy and safety.	Given the warnings associated with some of these newer classes, it seems less likely that they'll be embraced for use in pregnancy.	N/A	N/A	N/A

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
I-Study Needs/ Challenges	Populations • Racial distribution of study participants Study Designs • RCTs • Observational (population-based) studies to observe patterns of drug utilization and its effect on maternal and fetal outcomes Analyses • Intention-to-treat analyses • Specific analyses by race, socioeconomic status, education level, and access-to-care (e.g. insurance status, language, transportation) Outcomes • Consistent ascertainment, definition, and measurement of outcomes (across studies) Short-Term – Neonatal and maternal hypoglycemia, birth trauma, perineal tears, postpartum hemorrhage, etc. Long-Term – Infantile growth trajectory, neonatal & maternal metabolic alterations & adiposity, etc. Other Issues • Optimal glucose thresholds for initiating therapy? • Consistent reporting of baseline glucose tolerance levels	-	-	We need to standardize how we diagnose GDM.	In analyses, add in previous GDM and family history of diabetes.	Pharmacologic, pharmacokinetic studies, animal models to look at potential fetal effects and long-term offspring effects are of great importance here. Furthermore, many of the outcomes of interest are rare and would require large sample sizes (shoulder dystocia, for example). Long-term followup studies are expensive and labor intensive. Long-term studies must include more than the neonatal and early childhood periods. Older studies on obesity show that macrosomic offspring experience "catch-down" growth in the first few years, only to re-emerge as obese in late childhood and early teens.	-	Should development of type 2 diabetes mellitus in the mother postpartum be added as a long-term outcome?	Sample size, especially for subgroup comparisons. Not clear if neonatal development should be measured as well. Consistency of timing of followup postpartum for maternal and neonatal measures. Within treatment type heterogeneity. Inclusion of lifestyle interventions as a confounder.	Postpartum weight retention in women, and gestational weight gain after GDM diagnosis and treatment initiated.	Measuring adherence. The question of optimal glucose thresholds for initiating therapy could stand alone as a research question.	N/A	N/A	N/A

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
II-1	What is the effectiveness and safety of elective labor induction at 40 weeks compared to expectant management in women with gestational diabetes* with regard to the following maternal and neonatal outcomes? *Populations of Interest:• All women with gestational diabetes • Women with insulin-requiring (class A2) gestational diabetes• Obese women with gestational diabetes • Women with gestational diabetes with high estimated fetal weight (e.g. > 4000 or > 4500 grams)• Women with different parities • Women of various races/ethnicities	Question A: Is this research question clearly worded?	Yes	Yes	No. Again, reword outcomes to make them more patient centered. Actual question is fine.	Yes	Yes	Yes	Yes	No. Women treated with oral agents.	No. "What are the effectiveness and safety...". Same comments about neonatal outcomes. Is elective induction at 39 weeks specifically excluded from research interest? Why?	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	7	8	8	9	8	6	8	7	7.8	6	9
		Question C: Rationale for above rating	Consider cost-effectiveness as well as effectiveness.	-	Need to know the extent of influence mode and timing of delivery has on outcome.	This is a very pressing question. I would probably also select 39 weeks as a potential time for elective induction to be evaluated, since a number of epidemiologic studies have demonstrated significantly greater neonatal morbidity when elective induction of repeat cesarean section is performed before 39 weeks.	There is a paucity of evidence regarding this issue. In contrast, there is an abundance of opinion.	-	Definition of what cesarean section indication is critical. Consistency and complete measurement has been very problematic.	-	-	N/A	N/A	N/A

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
II-2	What is the effectiveness and safety of elective cesarean delivery at 40 weeks compared to expectant management in women with gestational diabetes* with regard to the following maternal and neonatal outcomes? *Populations of Interest: • All women with gestational diabetes • Women with insulin-requiring (class A2) gestational diabetes • Obese women with gestational diabetes • Women with gestational diabetes with high estimated fetal weight (e.g. > 4000 or > 4500 grams) • Women with different parities • Women of various races/ethnicities	Question A: Is this research question clearly worded?	No. As mentioned for previous section, consider cost-effectiveness as well as effective-ness.	Yes	Yes	Yes	Yes	Yes	No. Without clear specifications of what elective cesarean section is/is not, this isn't clear.	Same comment as previous.	No. Same comments as for Question II-1.	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	7	8	4	9	8	7	7	7	7.3	4	9
		Question C: Rationale for above rating	-	-	Same as previous question.	With the exception of the need to have an appropriate definition of macrosomia to trigger cesarean section without labor, I don't think this is a very compelling question. I believe that randomized trials would be necessary to answer the question, and randomization to elective cesarean section in patients whose EFW is not large would be very problematic.	See prior comment.	-	-	-	-	N/A	N/A	N/A

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
II-Study Needs/ Challenges	Study Designs • Observational studies with adjustment for factors that affect labor management (e.g. SES) • RCTs/ observational studies to address cesarean delivery based on EFW or gestational age Analyses • Stratified analyses (if diet-controlled and insulin-controlled women with gestational diabetes are included) • Stratified analyses by race/ethnicity • Stratified analyses by prior delivery history	-	-	What about stratification by glucose control?	Also stratify by previous GDM.	-	-	-	Measurement of indication for cesarean section is a major challenge. Stratif by oral agent subgroups with GDM.	Stratify by oral agent subgroups with GDM.	-	N/A	N/A	N/A
III-1	What is the evidence that maternal health behaviors (e.g. breastfeeding, physical activity, diet) are associated with the risk of developing type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	Physical activity and diet during pregnancy vs. after pregnancy, or no distinction?	Yes	Yes	Yes	Yes	Yes	Yes	No. Lactation duration and intensity.	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	9	9	8	8	7	7	9	7	8.1	7	9

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question C: Rationale for above rating	-	With the epidemic of diabetes we are facing, prevention is critical.	Resulting information will help target women who require close observation postpartum.	We are experiencing an epidemic of type 2 diabetes and if we can better predict who will get it we may be better able to intervene.	-	-	-	Only one study examined the risk of type 2 diabetes mellitus among women with a history of GDM, and the study found no association in the retrospective analysis of white women. This study relied on self-report of type 2 diabetes mellitus, and did not obtain and biochemical measurements of glycemia before or after lactation.	-	N/A	N/A	N/A
III-2	What is the evidence that maternal psychosocial factors (e.g. mood disorders, eating disorders, stress) are associated with the risk of developing type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	Yes	Yes	No. Would you also add in issues such as domestic violence and substance use disorders?	Yes	Yes	No. Should stress be defined as psychosocial stress? I assume you aren't referring to stress associated with heavy physical work.	No. Substance use disorders (not clear what is included in mood disorders).	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	7	8	5	4	6	6	4	5	6.0	4	9
		Question C: Rationale for above rating	Very important and totally understudied.	-	Again, information helps to both target women at high risk and provide medical rationale for addressing situations that are otherwise overlooked as not the role of the medical provider.	I'd include with previous research question.	-	-	-	These factors are likely mediated by weight gain and changes in lifestyle behaviors that have adverse effects on glucose tolerance.	-	N/A	N/A	N/A

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
III-3	What is the evidence that maternal metabolic measures (e.g. fasting insulin levels, OGTT measures, HPA axis stress (subclinical hypercortisolism)) are associated with the risk of developing type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	No. These are very broad categories. Consider subcategorizing maternal metabolic measures into more specific categories, such as adipocytokines; stress, and perhaps listing other categories-chemokines, vitamin D, etc.	Yes	Yes	Yes	Yes	Yes	No. Not clear if metabolic measures are during pregnancy or postpartum.	-	-	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	7	9	5	7	6	5	3	8	6.6	3	9
		Question C: Rationale for above rating	Risk stratification markers are needed.	Knowledge of risk factors will hopefully identify potential strategies for prevention.	`Would it also be important to add in the timing of those measures: 6 weeks postpartum, 6 months postpartum??	Include with previous two questions.	-	-	-	Many studies have already been published which demonstrate that these risk factors predict type 2 diabetes mellitus after GDM pregnancy.	-	N/A	N/A	N/A
III-4	What is the evidence that pre-pregnancy, antenatal, and postpartum co-morbid conditions (e.g. obesity, hypertension, hypercholesterolemia) are associated with the risk of developing type 2 diabetes or glucose intolerance/ prediabetes following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	No. Consider listing pre-conception biomarkers, such as the ones listed in the previous question.	Yes	Yes	Yes	Yes	Yes	No. Need to clarify what conditions are prepregnancy, pregnancy induced, and when they are measured postpartum. Without these clarifications there is too much ambiguity.	No. Pre-pregnancy cardiometabolic risk factors (atherogenic lipid profiles, pre-diabetes, abdominal obesity).	-	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	7	8	5	9	6	5	9	8	7.3	5	9

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question C: Rationale for above rating	Monique Hedderson has had an R01 (score 8th percentile) which is examining pre-pregnancy factors for prediction of GDM. I am not sure if powered or funded for longer followup postpartum.	As with previous questions ...	-	These are all variables that could be included in the first question in this series. All are important to study.	-	-	This is somewhat messy given pre, ante, and postnatal are all included at once. It will be difficult to determine what the critical period is. If it is pregnancy induced (hypertension) or not will be important to differentiate. The question is important/interesting, just ambiguous.	Pre-pregnancy metabolic risk factors should be identified to screen and identify women during the interconceptual period who are at risk for GDM in a future pregnancy. Atherogenic lipid profiles include low HDL and elevated triglycerides and not simply hyper-cholesterolemia.	-	N/A	N/A	N/A
III-5	What is the evidence that contraceptive method (e.g. depo-provera) is associated with the risk of developing type 2 diabetes or glucose intolerance/ prediabetes following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	Yes	Yes	No. Add in hormonal contraceptive method used within x months of pregnancy. Women often use multiple methods, especially those who become pregnant.	Yes	Yes	Yes	No. What kind of contraceptive method? Is the question addressing classes of contraceptives ? This question lacks specificity.	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	3	7	3	9	3	5	6	2	6	4.9	2	9
		Question C: Rationale for above rating	Studied in small populations already.	-	Interesting question, however the timing of the contraceptive method to the pregnancy and the consistency of use of one method would make this a difficult research question.	We desperately need more information about appropriate contraception for former GDMs … all future pregnancies should be planned pregnancies for them.	-	-	-	The glucose intolerance associated with these methods is well-known, and not recommended in women with previous GDM.	-	N/A	N/A	N/A

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
III-6	What is the evidence that the inter-conception interval is associated with the risk of developing type 2 diabetes or glucose intolerance/prediabetes following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	Yes	No. Please identify factors under consideration, i.e. weight gain, behaviors, physical activity. Is that what was intended by the question?	No. Add an amount of time such as under 12 months after inter-conception interval and compare that to women with an inter-conception interval over 12 months.	Yes	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	5	7	7	5	7	5	4	5	5	5.6	4	7
		Question C: Rationale for above rating	This is interesting although it might be very difficult to study since interval between pregnancies confounded by other factors (access to contracep-tion; socio-economic status, etc.).	-	-	Interesting question, just not as high priority for me as some of the previous ones, since decisions about family planning are tough to influence even if we know the answer.	I'm uncertain whether this has been studied before but it is an interesting question.	-	It is unclear if the first or second pregnancy is the GDM index pregnancy. The potential for changing pregnancy interval as an intervention is rather distal and there are likely more effective interventions that address other factors.	Somewhat important but primarily related to postpartum weight retention.	-	N/A	N/A	N/A
III-7	What is the evidence that family history, gene mutations, genotypes, gene-environment interactions, epigenetic modifications, or other biomarkers are associated with the risk of developing type 2 diabetes among women with gestational diabetes? Are there differences in these associations by race or ethnic group?	Question A: Is this research question clearly worded?	No. Very broad-consider subdividing into sub-categories.	Yes	Yes	Yes	Yes	Yes	No. How is this going to be measured? This covers just about everything one might think of as far as genetics goes. It would be better to focus on a subset of the list where there could be some possibility of useful interventions being developed.	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	9	9	8	9	5	3	7	8	7.4	3	9

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question C: Rationale for above rating	-	-	Data to identify at risk women and target resources.	These are critical questions in dealing with the epidemic of type 2 diabetes.	A straight-forward and objective study.	-	The question as written will not provide an evidence base that can be used. Because the question is so broad nothing can be done with the findings—lack of comparability.	Specific genetic markers among women with GDM are not well characterized.	-	N/A	N/A	N/A
III-8	What is the comparative effectiveness of various lifestyle interventions for prevention of type 2 diabetes mellitus, glucose intolerance, and obesity in women with a history of gestational diabetes?	Question A: Is this research question clearly worded?	Yes	Yes	No. Define lifestyle interventions.	Yes	Yes	No. Some examples of lifestyle interventions would be helpful, i.e., diet, physical activity, smoking.	No. Without some clear definition of lifestyle interventions findings from such studies will not be useful to the larger population.	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	9	8	6	8	7	6	8	8	7.7	6	9
		Question C: Rationale for above rating	Not much out there! It might be hard to compare comparative effectiveness. This kind of info is limited for lifestyle interventions outside of pregnancy--but could consider value of face-to-face vs. internet vs. telephone, etc. I think Cochrane has done something like this, at least for diabetes treatment or weight management.	Prevention is critical.	Strong evidence of lifestyle intervention as a preventive measure for diabetes may improve resources for promotion.	The results of the large NIH sponsored study on prevention of type 2 diabetes shed some light on this, but further data specific to women with history of GDM would be helpful.	Very important for prevention.	-	Comparative effectiveness research should include lifestyle and pharmacological interventions. Not clear why this is limited to just lifestyle.	Few RCTs have been conducted in this population to assess appropriate timing for the intervention.	-	N/A	N/A	N/A

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
III-9	What is the comparative effectiveness of various innovative strategies and technologies to disseminate educational materials on prevention of type 2 diabetes mellitus and glucose intolerance to women with a history of gestational diabetes?"	Question A: Is this research question clearly worded?	Yes	Yes	Yes	Yes	Yes	No. Some examples of innovative strategies and technologies would be helpful.	No. Various innovative strategies is too vague.	Yes	No. Remove the word 'innovative', as one wouldn't want to exclude the possibility that a non-innovative (low cost, sustainable) strategy could favorably compare to an innovative (high cost, nonsustainable) strategy. Any strategy should be eligible for testing.	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	1	9	9	8	5	6	6	8	7	6.6	1	9
		Question C: Rationale for above rating	Doesn't really matter how you disseminate materials if they don't work. It has been shown that knowledge is not the primary barrier.	-	Wonderful. These women are consumers at this point, not patients and make their own choices based on identified needs. Reaching them with clear compelling messages that trigger action is key.	Obviously if successful strategies can be found it would be very helpful.	-	-	Definition of strategies is too vague for multiple researchers to take this question on and have an impact on the body of evidence. This is a high risk population so it has the potential to be important.	-	-	N/A	N/A	N/A

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
III-Study Needs/ Challenges	Populations • Consistency in race/ethnicity (across studies) • Need for recruitment at time of gestational diabetes diagnosis Study Designs • Observational studies with ascertainment of covariates and confounders • Random or purposeful sampling for more representative samples Analyses • Multivariate analyses for pooled estimates of risk Other Issues • Consistency in risk factors and their definitions (across studies) • Focus on specific categories of anthropometry (e.g. BMI, weight) and reproductive factors (e.g. parity)	-	-	As before, consistency in diagnosis.	-	RCTs would be helpful and important in evaluating prevention strategies.	-	Under other issues, one should also consider age (i.e., the older pregnant woman).	Timing of measurement for factors (prepregnancy, during or postpartum, and then when post partum). Definition of interventions being compared and if they are evidence based or not.	Lactation support needs to be evaluated in the intervention trials, and when to intervene during the postpartum period.	-	N/A	N/A	N/A
IV-1	What are the performance characteristics (sensitivity, specificity, and reproducibility) of a single fasting blood glucose test compared to the full 2-hour 75-gm OGTT in screening for type 2 diabetes following a pregnancy with gestational diabetes? Does the accuracy of the fasting blood glucose test compared to the full 2-hour 75-gm OGTT vary with the postpartum testing interval in screening for type 2 diabetes following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	No.Already studied--as long as OGTT is considered as gold standard, FBG will always perform more poorly than OGTT. To resolve this question, we need to have long-term data on complications.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question B: Likely clinical benefit? (scale=1 to 9)	1	9	9	4	8	7	6	4	5	5.9	1	9
		Question C: Rationale	Should consider HgbA1c as well, since ADA has recommended that as a screening test.	Women are not receiving their postpartum followup. Need to explore similar means of diagnosis to increase compliance.	Extremely important. Difficult to capture women postpartum for testing, particularly a 2 hour OGTT however women are more amenable to a quick fasting blood sugar. Learning something about the correlation could be critical.	Given the recent ADA recommendations for HgbA1c to diagnose diabetes mellitus (but not so good for prediabetes), and the already demonstrated inadequacy of fasting glucose determinations to find cases of impaired glucose tolerance, and the importance of diagnosing prediabetes in these women who are "at risk" for another pregnancy in the near future, I believe that this issue is not at the top of the list. If it is to be done, HgbA1c levels should also be evaluated for their efficiency at diagnosis.	-	-	-	Evidence shows that a high proportion of women with glucose intolerance after GDM are missed if fasting only is used.	Research on the testing interval seems more important than research comparing tests.	N/A	N/A	N/A
IV-2	What are the performance characteristics (sensitivity, specificity, and reproducibility) of the HgBA1c test compared to the 2-hour 75-gm OGTT in screening for type 2 diabetes following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	No. I see you already thought of this! While less studied, similar limitations- HgbA1c will do more poorly. But I think it is important.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No. Add testing interval component.	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	9	9	5	8	7	5	9	5	7.3	5	9

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question C: Rationale for above rating	See above and previous comment.	As with previous question....	The necessary timing of the HgbA1c at over 4 months postpartum makes it close to impossible for women without private insurance to be tested as state Medicaid runs out 90 days postpartum. The 2 hr OGTT can be preformed at 6 weeks postpartum.	By current standards, neither HgbA1c nor the 75 gram OGTT are screening tests...they are diagnostic tests. Again, I believe that the diagnosis of prediabetes is particularly important for this group of subjects.	Might make diagnosis easier.	-	-	HgbA1c test compared with the 2 hr OGTT is an extremely important comparison.	-	N/A	N/A	N/A
IV-3	What is the comparative effectiveness of various strategies or interventions to improve clinician compliance with postpartum screening guidelines for type 2 diabetes mellitus and glucose intolerance in women with a history of gestational diabetes?	Question A: Is this research question clearly worded?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	9	9	7	5	7	8	9	7	7.8	5	9
		Question C: Rationale for above rating	Definitely needed. Might help if you give some examples--automated reminders, mailings, patient case-management, etc.	Sound strategies are needed on both sides.	-	This process is tedious and long since we can't stop and then resume at a later time. I am getting grouchy and thus less likely to describe my rationale. There must be a way to enhance the program so we can partially complete the task and then go back to it.	The strategies would need to be very specific.	-	This is a practice pattern that has the potential to improve the health of women and is inconsistently done. Getting this done and implementing appropriate interventions has the potential to improve health.	-	-	N/A	N/A	N/A

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
IV-4	What is the comparative effectiveness of various innovative strategies or technologies (e.g. electronic health records) to track postpartum screening for the development of type 2 diabetes mellitus and glucose intolerance in women with a history of gestational diabetes?	Question A: Is this research question clearly worded?	No. Anything else besides electronic health records? Need something to compare them to.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	5	7	9	4	4	7	8	9	6	6.6	4	9
		Question C: Rationale for above rating	-	-	-	-	-	-	-	Kaiser Permanente has already implemented the electronic medical record nationwide which tracks laboratory screening tests of members.	-	N/A	N/A	N/A
IV-Study Needs/ Challenges	Populations• Consistency in terms of race/ethnicity and future diabetes risk (family history of type 2 diabetes, prior gestational diabetes, etc.) (across studies)• Consistent protocols for recruitment of participants (across studies)Other Issues• Do the screening test and interval of followup vary based on maternal risk factors other than gestational diabetes?	-	-	-	Add in insurance factors and family stability - a new baby may mean moving and poor means of followup.	-	-	-	Timing of first screen postpartum.	-	-	N/A	N/A	N/A

Appendix D. Results and comments from external stakeholders (Delphi round 1) (continued)

Research Question Number	Research Question	Sub-question(s)	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
Additional	-	Do you have any general comments?	Very nice formulation of issues! Comments were minor editorial. Consider placental factors, inclusion of lipids particularly free fatty acids.	-	Thanks for involving me in the survey.	Too long to do all in one session. My later answers are undoubtedly less thoughtful than earlier ones.	-	-	-	-	Well done! Thanks very much for asking me to participate.	N/A	N/A	N/A

Abbreviations: ADA=American Diabetes Association, FBG=fasting blood glucose, GDM=gestational diabetes mellitus, HDL=high density lipoprotein, OGTT=oral glucose tolerance test, RCT=randomized controlled trial.

Appendix E. Results and Comments from External Stakeholders (Delphi Round 2)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
I-3	What are the comparative effectiveness and safety of various insulin regimens in terms of type/duration, dosing, and frequency of administration in the treatment of gestational diabetes with regard to the following short- and long-term maternal outcomes, neonatal outcomes , and long-term offspring outcomes?	Question A: Is this research question clearly worded?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No. The distinction between short- and long-term outcomes is unclear in this context. I would recommend more specific terms that could include "pregnancy" or "perinatal" outcomes and "postpartum" maternal outcomes for women instead of short- and long-term. The term long-term would encompass the years after delivery and incident type 2 diabetes years in mid life or later rather than solely during the postpartum period. I also recommend that a postpartum maternal outcome is obesity in addition to postpartum weight retention.	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	7	6	9	7	7	8	7	8	7	7.3	6	9
		Question C: Rationale for above rating	-	-	-	-	-	-	-	The last phrase is too long and awkward, "following short- and long-term maternal outcomes, neonatal outcomes , and long-term offspring outcomes". I suggest that you state, "perinatal and postpartum maternal outcomes, and the neonatal and childhood offspring outcomes".	-	N/A	N/A	N/A

Appendix E. Results and comments from external stakeholders (Delphi round 2) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
I-4	What are the effectiveness and safety of other hypoglycemic drug classes (e.g. thiazolidinediones, DPP-4 inhibitors, GLP1 agonists, meglitinides) compared to any insulin or other hypoglycemic drugs in the treatment of gestational diabetes with regard to the following short- and long-term maternal outcomes, neonatal outcomes, and long-term offspring outcomes?	Question A: Is this research question clearly worded?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No. Same comments as the previous	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	8	3	9	7	7	8	6	8	4	6.7	3	9
		Question C: Rationale for above rating	-	Should address the safety and effectiveness of the two oral agents being clinically utilized before we move forward with other agents.	-	-	-	-	-	Instead of this long wording "short- and long-term maternal outcomes, neonatal outcomes, and long-term offspring outcomes", I suggest that you state, I suggest that you state, "perinatal and postpartum maternal outcomes, and the neonatal and childhood offspring outcomes".	-	N/A	N/A	N/A
III-2	What is the evidence that maternal psychosocial factors (e.g. mood disorders, substance use disorders, eating disorders, stress) are associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	Yes	Yes	No. Make it "tobacco and substance use disorders".	Yes	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	8	6	8	4	4	6	8	5	5	6.0	4	8

Appendix E. Results and comments from external stakeholders (Delphi round 2) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question C: Rationale for above rating	Understudied and probably important as these are common factors. In an unrelated comment, the ranking of "3 most important" outcomes is somewhat artificial--of course you'd study more. I also have suspicion that "fundable" outcomes ("hard" outcomes like cesarean delivery, other "medical" complications) might differ from other outcomes that might actually have greater influence upon complication rates in the 'real world', such as patient preferences. If a patient doesn't want 3 time daily injectable insulin, would a slight decrease in cesarean section rates be enough to convince her.	No comments	It is unclear whether it is the risk factor itself or the maternal response to the risk factor that increases the incidence of type 2 diabetes following GDM and that helps decide targeting patient's at risk and patient management.	I'm not sure this question is related only to pregnancy; same question could be posed regarding the development of diabetes in males, or females who have never been pregnant.	No comments	No comments	No comments	No comments	No comments	N/A	N/A	N/A

Appendix E. Results and comments from external stakeholders (Delphi round 2) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
III-3	What is the evidence that maternal metabolic measures (e.g. fasting insulin levels, OGTT measures, HPA axis stress (subclinical hypercortisolism)) are associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	No. Maternal metabolic measures during pregnancy or after pregnancy? If after pregnancy, is there any reason to think they differ from women in general? We know that insulin and glucose are associated with diabetes risk and are in the pathway to diabetes risk. HPA stress axis is important but sticks out as the more novel factor that is in a different category than the others. Consider separating HPA out, or dumping fasting insulin and glucose levels.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	7	7	8	8	8	7	8	6	7	7.3	6	8
		Question C: Rationale for above rating	-	-	Needed to determine who is at most risk and in need of close followup.	-	-	-	-	-	-	N/A	N/A	N/A

Appendix E. Results and comments from external stakeholders (Delphi round 2) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
III-4	What is the evidence that co-morbid conditions (e.g. advanced maternal age, obesity, hypertension, hypercholesterolemia) are associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	No. How about adding, "assisted reproduction" or "fertility procedures" or some such?	Yes	Yes	Yes	Yes	Yes	Yes	No. I suggest that these time periods be specified because it is important that each is specifically recognized as important. Currently these conditions are not being advocated prior to pregnancy. Other conditions before pregnancy that are not recognized include hyperglycemia or impaired fasting glucose and glucose intolerance.	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	7	6	8	8	7	6	8	8	7.4	6	9
		Question C: Rationale for above rating	The order in which these factors unfold is important. this was a big focus of the GDM Pasadena conference this year--more specifically, many of the bench presentations focused on lipids and how this predicts future glucose tolerance, and many of the other presentations focused on how obesity predicted perinatal complications.	-	For followup, women with comorbid conditions are more likely to receive postpartum followup and less likely to slip through the cracks.	-	-	-	-	It is important to specify that each period is important for assessing risk. I do not agree with the suggestion to make this generic. We do not want the focus to be on the postpartum period only. Currently, these comorbid conditions are not measured routinely before pregnancy, but should be assessed during the intercurrent period.	-	N/A	N/A	N/A

Appendix E. Results and comments from external stakeholders (Delphi round 2) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
III-5	What is the evidence that contraceptive method (e.g. depo-provera) is associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	No. Depo is not a good example. You may want to use - hormonal contraception or combined contraception pills.	Yes	Yes	Yes	Yes	Yes	Yes	No. Add to the list of diseases at risk of developing... "other cardiometabolic diseases (i.e. the metabolic syndrome)".	No. Shouldn't use trade name for drug, substitute depot medroxypregesterone acetate.	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	8	7	8	8	3	5	5	4	3	5.7	3	8
		Question C: Rationale for above rating	-	Moving forward the prevention of diabetes is critical. If true, solution is relatively straight forward.	-	-	-	-	-	-	-	N/A	N/A	N/A
III-6	What is the evidence that the inter-conception interval is associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No. Add to the list of diseases to risk of developing ...."other cardiometabolic diseases (i.e. the metabolic syndrome)".	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	8	5	6	6	2	5	6	4	6	5.3	2	8
		Question C: Rationale for above rating	-	-	-	-	-	-	-	-	-	N/A	N/A	N/A

Appendix E. Results and comments from external stakeholders (Delphi round 2) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
III-9	What is the comparative effectiveness of various educational and behavioral supportive strategies (e.g. patient education about diabetes risk, lactation support) to prevent type 2 diabetes mellitus and glucose intolerance/impaired fasting glucose in women with a history of gestational diabetes?	Question A: Is this research question clearly worded?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No. Add to the list of diseases to prevent... "other cardiometabolic diseases (i.e. the metabolic syndrome)".	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	8	8	8	2	7	7	9	8	7.3	2	9
		Question C: Rationale for above rating	Such a demonstration of effectiveness might lead to insurance companies reimbursing for these services.	-	-	-	A more important question would be how they affect fasting and postprandial glucose levels.	-	-	Delete the word "supportive" strategies, and replace with "change" strategies. In the list, I would include: (e.g., disease risk, dietary, physical activity, and lactation). These are appropriate to include since the word "innovative" was removed. Also, if the other questions expand disease risk beyond type 2 diabetes mellitus then should consider adding those endpoints.	-	N/A	N/A	N/A

Appendix E. Results and comments from external stakeholders (Delphi round 2) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
IV-1	What are the performance characteristics (sensitivity, specificity, and reproducibility) of a single fasting blood glucose test compared to the full 2-hour 75-gm OGTT in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes? Does the accuracy of the fasting blood glucose test compared to the full 2-hour 75-gm OGTT vary with the postpartum testing interval in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	No. Consider using the terms reliability, precision, and accuracy rather than sensitivity and specificity--we already know that sensitivity is lower with FBG.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	6	8	9	6	7	7	6	4	6	6.6	4	9

Appendix E. Results and comments from external stakeholders (Delphi round 2) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question C: Rationale for above rating	Is there a reason that the performance characteristics of these tests would vary from the performance characteristics of other populations? Age might make an impact, but I'm not sure about others. If not, the impact of answering this question (FBG vs. OGTT) might be limited--we already know that FBG is less sensitive and more reliable as well as more precise.	-	Because pregnancy Medicaid coverage ends at 8 to 9 weeks postpartum and because low income women are a high risk population for GDM and type 2 diabetes, the availability of accurate testing at the earliest possible date postpartum is of utmost importance.	I'm not sure this question should be specific to former GDMs. It is a very important question in general for testing for prediabetes and diabetes.	-	-	-	-	-	N/A	N/A	N/A
IV-2	What are the performance characteristics (sensitivity, specificity, and reproducibility) of the HgBA1c test compared to the 2-hour 75-gm OGTT in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No. Should include the followup question as in IV-1.	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	8	9	7	8	8	6	9	7	7.9	6	9
		Question C: Rationale for above rating	This is particularly important with the new ADA recommendations.	-	Same comment as previous question. An accurate test that can not be done because the woman is no longer insured is unacceptable.	Not specific to former GDMs. The question of an appropriate HgbA1c cutoff for prediabetes is important for all testing situations.	-	-	-	-	-	N/A	N/A	N/A

Appendix E. Results and comments from external stakeholders (Delphi round 2) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
IV-4	What is the comparative effectiveness of different health information technology interventions to track postpartum screening for the development of type 2 diabetes mellitus and glucose intolerance/impaired fasting glucose in women with a history of gestational diabetes?	Question A: Is this research question clearly worded?	Yes	Yes	No. After "health information technology interventions" mention "including social media".	Yes	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	5	6	9	7	2	7	7	6	7	6.2	2	9
		Question C: Rationale for above rating	-	-	Use of text messaging or 'Facebook' for engaging, reminding, and tracking women who had GDM needs to be assessed.	-	-	-	-	-	-	N/A	N/A	N/A
Additional	-	Do you have any general comments?	I mentioned earlier that choosing 3 outcomes feels too restrictive.	-	-	The length of this one was much more user-friendly and I don't object to having to do it all in one sitting.	-	-	-	-	-	N/A	N/A	N/A

Abbreviations: ADA=American Diabetes Association, FBG=fasting blood glucose, GDM=gestational diabetes mellitus, HPA=hypothalamo-pituitary-adrenal, OGTT=oral glucose tolerance test.

Appendix F. Results and Comments from External Stakeholders (Delphi Round 3)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
I-4	What are the effectiveness and safety of other hypoglycemic drug classes (e.g. thiazolidinediones, DPP-4 inhibitors, GLP-1 agonists, meglitinides) compared to any insulin or other hypoglycemic drugs in the treatment of gestational diabetes with regard to the following short- and long-term maternal outcomes, neonatal outcomes , and long-term offspring outcomes?	Question A: Is this research question clearly worded?	Yes	Yes	No. The use of the term "or other hypoglycemic drugs" in the second line is confusing because you are comparing with "other hypoglycemic drug classes".	Yes	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	9	4	7	5	8	7	7	7	8	6.9	4	9
		Question C: Rationale for above rating	-	I am still not clear why we would want to move forward with other oral agents when the long-term impact of metformin and glyburide on maternal and neonatal outcomes is still not known. Should focus on agents currently being utilized clinically.	-	I am relatively neutral because I believe that so much needs to be done with glyburide, metformin, and alphaglucosidase inhibitors which are already being used that I hate to see attention and resources diverted from them. On the other hand, if the meds noted do not turn out to be safe or effective, then we should move on to the others after animal studies better address the issue of transplacental transfer and fetal exposure.	-	-	-	-	-	N/A	N/A	N/A

Appendix F. Results and comments from external stakeholders (Delphi round 3) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
III-2	What is the evidence that maternal psychosocial factors (e.g. mood disorders, substance use disorders, eating disorders, stress) are associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	Yes	Yes	Yes	Yes	Yes	Yes	No. The suggested list of psychosocial factors is really psychological rather than social. The other problem is that there is no time frame—within five years, 50 years? Needs clarification.	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	7	6	7	5	8	6	5	5	5	6.0	5	8
		Question C: Rationale for above rating	-	-	Understanding the relationship between psychosocial and metabolic factors may further guide prevention strategies and identify those at risk.	Worthwhile questions but not directly related to GDM...applicable to diabetes even without history of GDM.	-	-	-	-	-	N/A	N/A	N/A
III-5	What is the evidence that contraceptive method (e.g. progestin-only) is associated with the risk of developing type 2 diabetes or glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	No. Change to "progestin-only oral contraceptive pills" or "progestin-implant" OR "progestin injection".	Yes	Yes	Yes	Yes	Yes	No. See below.	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	8	7	6	8	5	5	5	5	4	5.9	4	8

Appendix F. Results and comments from external stakeholders (Delphi round 3) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question C: Rationale for above rating	-	Prevention of type 2 diabetes is of critical importance.	-	-	-	-	Timeframe is an issue here---developing diabetes in what time frame, after what duration of using the contraceptive method (ever use, after at least 5 years of use), it appears that it is implied as hormonal contraceptives is the contraceptive of interest. There are other contraceptive methods in the generic definition.	-	-	N/A	N/A	N/A
IV-1	What are the performance characteristics (sensitivity, specificity, and reproducibility) of a single fasting blood glucose test compared to the full 2-hour 75-gm OGTT in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes? Does the accuracy of the fasting blood glucose test compared to the full 2-hour 75-gm OGTT vary with the postpartum testing interval in screening for type 2 diabetes and glucose intolerance/impaired fasting glucose following a pregnancy with gestational diabetes?	Question A: Is this research question clearly worded?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A

Appendix F. Results and comments from external stakeholders (Delphi round 3) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question B: Likely clinical benefit? (scale=1 to 9)	1	7	9	7	7	7	6	8	8	6.7	1	9
		Question C: Rationale for above rating	I think the question has been answered.	The data indicate that the majority of women with a history of GDM do not get followup testing. If performance of FBG was found to be adequate, it could potentially substantially increase the number of women who are tested.	This is very clear and extremely important to determine in light of the difficulty to "capture" postpartum women for a 2 hr OGTT and also to determine the earliest an accurate postpartum screening can be done and followup initiated prior the loss of insurance for many women on Medicaid.	The first of the two questions is not specific to GDM.	-	-	-	-	-	N/A	N/A	N/A
IV-4	What is the comparative effectiveness of different health information technology interventions to track postpartum screening for the development of type 2 diabetes mellitus and glucose intolerance/impaired fasting glucose in women with a history of gestational diabetes?	Question A: Is this research question clearly worded?	Yes	Yes	Yes	Yes	Yes	No. The linkage of different health information technology interventions with development of type 2 diabetes is unclear. Could a brief introductory rationale be provided? That would help identify what interventions could be considered.	Yes	Yes	Yes	N/A	N/A	N/A
		Question B: Likely clinical benefit? (scale=1 to 9)	5	6	8	7	5	5	6	7	3	5.8	3	8

Appendix F. Results and comments from external stakeholders (Delphi round 3) (continued)

Research Question Number	Research Question	Sub-question(s)	Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4	Stakeholder 5	Stakeholder 6	Stakeholder 7	Stakeholder 8	Stakeholder 9	Result Mean	Result Lower Range	Result Upper Range
		Question C: Rationale for above rating	It would be nice if examples of specific interventions could be provided, but understandable if they cannot. What is there besides an electronic medical record?	I feel the priority should be on getting women tested and then move to better tracking methods.	Essential for both internal office tracking and in reminding women of their need for followup. It would be great to have a texting program for postpartum reminders.	-	-	-	-	-	-	N/A	N/A	N/A
Additional	-	Do you have any general comments?	-	-	Thank you for the opportunity to participate. The process was painless and feedback educational.	-	-	-	-	-	-	N/A	N/A	N/A

Abbreviations: DPP-4=Dipeptidyl Peptidase-4, FBG=fasting blood glucose, GDM=gestational diabetes mellitus, GLP-1= Glucagon-Like Peptide-1, OGTT=oral glucose tolerance test.

Appendix G. Results from Step 8 (Evaluation of Entire Process by Evidence Report Authors and Local Stakeholders)

No.	Questions	Report Author 1	Report Author 2	Report Author 3	Report Author 4	Report Author 5	Local Stakeholder 1	Local Stakeholder 2	Local Stakeholder 3	Local Stakeholder 4	Local Stakeholder 5	Local Stakeholder 6
1	Did you have adequate information to effectively participate?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2	Which mode of participation would you have preferred?	Web-based feedback	Web-based feedback	Web-based feedback	Telephone	Web-based with an option to participate in in-person or telephone meetings during the process.	In-person	Web-based feedback	Web-based feedback	Web-based feedback	I thought the combination of a web-based and in-person feedback was good. Web-based survey was more time effective, but the in-person meeting allowed for richer conversation and feedback.	In-person
3	Looking at the final research questions, do you feel that we accomplished our objective?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Appendix G. Results from step 8 (Evaluation of entire process by evidence report authors and local stakeholders) (continued)

No.	Questions	Report Author 1	Report Author 2	Report Author 3	Report Author 4	Report Author 5	Local Stakeholder 1	Local Stakeholder 2	Local Stakeholder 3	Local Stakeholder 4	Local Stakeholder 5	Local Stakeholder 6
4	Are there any final research questions that you would re-phrase?	Yes. Research questions III-7. What is the evidence that family history, gene-environment interactions, and epigenetic alterations are associated with the risk of developing type 2 diabetes or glucose intolerance/ impaired fasting glucose among women with gestational diabetes?	Yes. I thought research questions IV-1 and IV-2 could be combined into one question (i.e., comparison of FBG vs HbA1c vs 2-hour PPG) but am ok with separate.	No	No	No	Yes. III-1. Suggest that since the question has a negative connotation as currently written and implies a directionality of effect of these behaviours (increased risk) just as this is implied in question below it. For example, if exercise lowers the risk of type 2 diabetes, I am unsure if this question would be the best.	No	No	No	Yes. Only would add birth defects as a neonatal outcome for questions I-1 through I-4.	No

Appendix G. Results from step 8 (Evaluation of entire process by evidence report authors and local stakeholders) (continued)

No.	Questions	Report Author 1	Report Author 2	Report Author 3	Report Author 4	Report Author 5	Local Stakeholder 1	Local Stakeholder 2	Local Stakeholder 3	Local Stakeholder 4	Local Stakeholder 5	Local Stakeholder 6
5	Was the composition of the local group of 6 stakeholders comprehensive? (i.e. was it adequate to include physicians (obstetricians and gynecologists), nutritionists/ dietitians, epidemiologists/ methodologists, and patient/ consumer representatives?)	Yes	Yes	Yes	No. One other group to consider input from would be neonatologists as they could provide insight into which neonatal outcomes are associated with the greatest moridibity/ mortality in the newborn period. Otherwise, the relevant stakeholders were included.	Yes	Yes	Yes	No. You should consider including patients either with GDM now or history of GDM for their perspective on the importance of these questions.	Yes	Yes	Yes

Appendix G. Results from step 8 (Evaluation of entire process by evidence report authors and local stakeholders) (continued)

No.	Questions	Report Author 1	Report Author 2	Report Author 3	Report Author 4	Report Author 5	Local Stakeholder 1	Local Stakeholder 2	Local Stakeholder 3	Local Stakeholder 4	Local Stakeholder 5	Local Stakeholder 6
6	Was the composition of the external group of 10 stakeholders comprehensive? (i.e. was it adequate to include physicians (obstetricians and gynecologists), nutritionists/ dietitians, epidemiologists/ methodologists, research funders, and patient/ consumer representatives?)	Yes	Yes	Yes	No. Similar to my answer to the previous question, I would have suggested including neonatology interests.	Yes	Yes	Yes	No. Same as above.	Yes	Yes	Yes
7	Could we have abbreviated our process in some way? (i.e. was it necessary to get feedback from the 2008 evidence report authors, local stakeholders, and external stakeholders?)	No	No	No	No	No	No	No	No	No	No	No

Appendix G. Results from step 8 (Evaluation of entire process by evidence report authors and local stakeholders) (continued)

No.	Questions	Report Author 1	Report Author 2	Report Author 3	Report Author 4	Report Author 5	Local Stakeholder 1	Local Stakeholder 2	Local Stakeholder 3	Local Stakeholder 4	Local Stakeholder 5	Local Stakeholder 6
8	Do you have any additional feedback on our process or on the final research questions?	No	Yes. It is probably too late to specify this but as I looked at research questions related to KQ-IV, I was struck that we didn't include any outcomes related to patient satisfaction/ adherence with postpartum testing. Anyways, sounds like it is too late to vet that question.	No	No	Yes. Would only recommend that if this process is implemented, the identification of research needs occurs at the time of the evidence report or immediately following preparation of the evidence report.	Yes. The space for typing comments on this survey is limited.	No	No	No	No	Yes. I was disppointed that the psychosocial and disparity factors were not part of final research questions. I wonder if the group had a greater percentage of consumer representatives that you would get additional questions to explore.

Abbreviations: FBG=fasting blood glucose, GDM=gestational diabetes mellitus, HbA1c=hemoglobin A1c, OGTT=oral glucose tolerance test, PPG=postpartum glucose.

Appendix H. Results from Step 8 (Evaluation of Entire Process by External Stakeholders)

No.	Questions	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9
1	Did you have adequate information to effectively participate?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2	Which mode of participation would you have preferred?	Web-based feedback	Web-based feedback	Web-based feedback	Web-based feedback that allows users to save a partially completed survey and come back at another time. This must be possible because I have participated in other web-based surveys that allow this.	Web-based feedback	Web-based feedback	Web-based feedback	Web-based feedback	Web-based feedback

Appendix H. Results from step 8 (evaluation of entire process by external stakeholders) (continued)

No.	Questions	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9
3	Looking at the final research questions, do you feel that we accomplished our objective?	I thought you accomplished your objective. It was possible to start from an even broader base—for example, the importance of genomics for GDM or consideration of diagnostic criteria in the context of strain on existing resources. I suppose the drawback of that would have been that it would have been too broad and no consensus would have been reached.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4	Are there any final research questions that you would rephrase?	Yes. The use of the term "accuracy" in comparing FBG and 2-hour OGTTs could be rephrased—FBGs will always be less sensitive and specific if OGTT is used as the gold standard. The only way to settle it is to make the criteria outcomes-based.	No	Yes. IV-1 and IV-2 are dependent on patient compliance even more than performance characteristics. The longer the test and the more distant from the delivery the poorer the compliance.	No	No	No	Yes. In research question I-4, the comparison is between other hypoglycemic drug classes to other hypoglycemic drugs. Perhaps this is clear to all except for me that drugs and drug classes are two different things.	No	No

Appendix H. Results from step 8 (evaluation of entire process by external stakeholders) (continued)

No.	Questions	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9
5	Was the composition of the local group of 6 stakeholders comprehensive? (i.e., was it adequate to include physicians (obstetricians and gynecologists), nutritionists/ dietitians, epidemiologists/ methodologists, and patient/ consumer representatives?)	Yes	No. I suggest including endocrinologists who care for women with diabetes.	Yes	Yes	Yes	Yes	No. Why weren't nephrologists included? This seems like a specialty that is missing.	I am unable to respond to this question.	Yes
6	Was the composition of the external group of 10 stakeholders comprehensive? (i.e., was it adequate to include physicians (obstetricians and gynecologists), nutritionists/ dietitians, epidemiologists/ methodologists, research funders, and patient/ consumer representatives?)	Yes	No. As with the internal group, would have included endcrinologists who care for pregnant women with diabetes.	Yes	No. Diabetologist input would have been a good idea.	Yes	Yes	No. Again, where were the nephrologists? Perhaps they provide care at a more advanced stage and aren't involved with primary prevention.	Yes	Yes
7	Could we have abbreviated our process in some way? (i.e., was it necessary to get feedback from the 2008 evidence report authors, local stakeholders, and external stakeholders?)	No	No	No	Yes. I'm not sure why a group of local stakeholders was necessary since the object of the process was nationally usable research questions. Perhaps it was used as a focus group, in which case the only reason for it to be local was convenience.	No	No	No	No	No

Appendix H. Results from step 8 (evaluation of entire process by external stakeholders) (continued)

No.	Questions	External Stakeholder 1	External Stakeholder 2	External Stakeholder 3	External Stakeholder 4	External Stakeholder 5	External Stakeholder 6	External Stakeholder 7	External Stakeholder 8	External Stakeholder 9
8	Do you have any additional feedback on our process or on the final research questions?	Yes. Nice job, obviously a lot of thought and hard work went into this.	No	No	Yes. Web-based instrument used should have allowed user to interrupt session and come back.	No	Yes. This process was very effective and efficient. Couldn't have been done better!!	Yes. Was consensus achieved after 3 rounds? In your final report you might consider documenting which ones didn't reach consensus. These are potentially the most controversial or debated in the field.	No	Yes. I enjoyed being a part of this and it seems to be an efficient yet nicely thorough method for decision-making. Thank you for including me. My only "complaint" would be the limitations of the online tool: not being able to save and return, and not being able to see any subsequent questions until all previous questions were answered. Having a simple way to download or print out the surveys for offline work/thought would have been helpful.

Abbreviations: FBG=fasting blood glucose, GDM=gestational diabetes mellitus, OGTT=oral glucose tolerance test.