Benefits and Harms of Treating Gestational Diabetes Mellitus: A Systematic Review and Meta-analysis for the U.S. Preventive Services Task Force and the National Institutes of Health Office of Medical Applications of Research

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

Background: Outcomes of treating gestational diabetes mellitus (GDM) are not well-established.

Purpose: To summarize evidence about the maternal and neonatal benefits and harms of treating GDM.

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

Study Selection: English-language randomized, controlled trials \( (n = 5) \) and cohort studies \( (n = 6) \) of women without known preexisting diabetes.

Data Extraction: One reviewer extracted data, and a second reviewer verified them. Two reviewers independently assessed methodological quality and evaluated strength of evidence for primary outcomes by using a Grading of Recommendations Assessment, Development and Evaluation approach.

Data Synthesis: All studies compared diet modification, glucose monitoring, and insulin as needed with no treatment. Women who were treated had more prenatal visits than those in control groups. Moderate evidence showed fewer cases of preeclampsia, shoulder dystocia, and macrosomia in the treated group. Evidence was insufficient for maternal weight gain and birth injury. Low evidence showed no difference between groups for neonatal hypoglycemia. Evidence was insufficient for long-term metabolic outcomes among offspring. No difference was found for cesarean delivery (low evidence), induction of labor (insufficient evidence), small-for-gestational-age neonates (moderate evidence), or admission to a neonatal intensive care unit (low evidence).

Limitations: Evidence is low or insufficient for many outcomes of greatest clinical importance. The strongest evidence supports reductions in intermediate outcomes; however, other factors (for example, maternal weight and gestational weight gain) may impart greater risk than GDM, particularly when glucose levels are modestly elevated.

Conclusion: Treating GDM results in less preeclampsia, shoulder dystocia, and macrosomia; however, current evidence does not show an effect on neonatal hypoglycemia or future poor metabolic outcomes. There is little evidence of short-term harm of treating GDM other than an increased demand for services.

Primary Funding Source: Agency for Healthcare Research and Quality.

For author affiliations, see end of text.
This article was published at www.annals.org on 28 May 2013.

Gestational diabetes mellitus (GDM) is defined as glucose intolerance first discovered in pregnancy. It predicts risk for overt diabetes in women. The more immediate risk for adverse outcomes of GDM in the mother and child is less well-established.

The prevalence of GDM ranges from 1.1% to 25.5% of pregnancies in the United States \( (1–3) \) and is influenced by diagnostic criteria and population characteristics, such as ethnicity. The incidence of this condition has increased over the past decades in parallel with the increase in rates of obesity and type 2 diabetes mellitus, and this trend is expected to continue.

Initial treatment of GDM involves diet modification, glucose monitoring, and moderate exercise. When dietary management does not achieve desired glucose control, insulin or oral antidiabetic medications may be used \( (4) \). Increased prenatal surveillance and changes in delivery management may also occur.

A report commissioned by the U.S. Preventive Services Task Force in 2008 found that treatment of women with mild GDM diagnosed after 24 weeks’ gestation improved maternal and neonatal health outcomes \( (5) \). Specifically, on the basis of 1 study, they found a reduction in “any serious perinatal complication,” which included death, shoulder dystocia, bone fracture, and nerve palsy \( (6) \). The number of events for many of the individual outcomes was extremely small, which did not provide adequate evidence to make conclusions for individual outcomes. The same study also found less depression and a trend to better quality of life 3 months after parturition and reduced maternal hypertension in the treated group \( (6) \).

Potential harms of GDM treatment may include small-for-gestational-age neonates; maternal stress; and additional costs, including those associated with laboratory testing as well as patient and clinician time \( (7) \). Anxiety of...
health care providers over the diagnosis could result in unnecessary or overly aggressive fetal and neonatal surveillance and delivery management. The purpose of this review is to evaluate whether treatment of GDM modifies outcomes of mothers and their offspring and whether it is associated with any harms.

**METHODS**

An a priori protocol was followed. Questions were developed by the Office of Medical Applications of Research and the U.S. Preventive Services Task Force. A technical expert panel that included representatives from both organizations provided content and methodological expertise. The full technical report is available at http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=1295&pageaction=displayproduct.

**Data Sources and Searches**

We searched for trials and cohort studies published in English from 1995 to May 2012 in MEDLINE (Ovid interface) (Appendix Table 1, available at www.annals.org), Ovid MEDLINE In-Process & Other Non-Indexed Citations, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Global Health, EMBASE, Pascal CINAHL Plus with Full Text (EBSCOhost), BIOSIS Previews (Web of Knowledge), Science Citation Index Expanded and Conference Proceedings Citation Index (both via Web of Science), PubMed, Latin American and Caribbean Health Science Literature, National Library of Medicine Gateway, and OCLC ProceedingsFirst and PapersFirst. We also searched trial registries and the Web sites of relevant professional associations and research groups for conference abstracts and proceedings between 2010 and 2012. We evaluated the reference lists of relevant reviews and included studies.

**Study Selection**

Two reviewers independently screened titles, keywords, and abstracts. We retrieved the full text for any study that was considered potentially relevant by at least 1 reviewer. Two reviewers independently assessed each full-text article by using a detailed form. We resolved disagreements through discussion. We included studies if they were randomized, controlled trials (RCTs) or non-RCTs or cohort studies; involved pregnant women with no known preexisting diabetes; compared any treatment of GDM with no treatment; and reported short- and long-term maternal, fetal, neonatal, and child outcomes that the technical panel deemed important.

**Data Extraction and Quality Assessment**

One reviewer extracted data by using a structured, electronic form, and a second reviewer checked the data for accuracy and completeness. Discrepancies were resolved through consensus. We extracted information on study characteristics, populations, interventions, outcomes, and results.

Two reviewers independently assessed the methodological quality of included studies and resolved disagreements through discussion. We used the Cochrane risk-of-bias tool to assess RCTs (8) and the Newcastle-Ottawa Scale to assess cohort studies (9).

**Data Synthesis and Analysis**

Two independent reviewers graded the strength of evidence by using the Evidence-based Practice Center Grading of Recommendations Assessment, Development and Evaluation approach (10). We resolved discrepancies by discussion. We assessed 4 major domains (risk of bias, consistency, directness, and precision) and summarized the overall strength of evidence for each outcome as high, moderate, or low. When no studies were available for an outcome or the evidence did not permit estimation of an effect, we rated strength of evidence as insufficient.

We described the results of studies qualitatively and in evidence tables. We performed meta-analyses when studies were sufficiently similar in terms of statistical homogeneity (that is, $I^2 \leq 75\%$). We used the Mantel–Haenszel method for relative risks and the inverse variance method for pooling mean differences.

We combined results by using the random-effects model (11). For dichotomous outcomes, we computed relative risk to estimate between-group differences. If no event was reported in 1 treatment group, a correction factor of 0.5 was added to each cell of the $2 \times 2$ table to obtain estimates of the relative risk.

For continuous variables, we calculated mean differences for individual studies. We reported all results with 95% CIs and used Review Manager Version 5.0 (The Cochrane Collaboration, Copenhagen, Denmark) to perform meta-analyses.

**Role of the Funding Source**

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

**RESULTS**

Of 14 428 citations, 5 RCTs (6, 12–15) and 6 retrospective cohort studies (16–21) met inclusion criteria (Appendix Figure 1, available at www.annals.org). All studies compared diet modification, glucose monitoring, and insulin as needed with standard care. Two studies had 2 associated publications reporting initial (6, 15) and longer-term (22, 23) outcomes. Diagnostic testing in all studies occurred at or after 24 weeks’ gestation (when reported).

Numerous glucose inclusion criteria were used, varying from screening positive on the 50-g glucose challenge with nondiagnostic oral glucose tolerance tests to meeting National Diabetes Data Group criteria for a diagnosis of...
GDM. The 2 largest RCTs used different glucose thresholds for entry in their trials: World Health Organization (6) and Carpenter–Coustan criteria with a fasting glucose level less than 5.3 mmol/L (95 mg/dL) (12); however, the mean glucose levels of women at study entry were similar between these 2 studies. Risk of bias was low for 1 trial (6), unclear for 3 trials (12–14), and high for 1 trial (15). All cohort studies were considered high quality, with overall scores of 7 to 9 on a 9-point scale.

Benefits of Treating GDM

The Table and Appendix Table 2 (available at www.annals.org) show results for maternal outcomes. Moderate evidence from 3 RCTs showed less preeclampsia with treatment (Appendix Figure 2, available at www.annals.org). In 2 of these trials, there was no difference between groups in gestational age at delivery. The strength of evidence for maternal weight gain was insufficient because of inconsistency across studies and imprecise effect estimates (Appendix Figure 3, available at www.annals.org). Two RCTs showed no difference (13, 15), whereas 2 large RCTs showed less weight gain with treatment (6, 12). Given the high body mass index (BMI) of the women studied, less gestational weight gain in the treatment group would be beneficial.

**Table. Strength of Evidence for Benefits and Harms of treating GDM**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Source</th>
<th>Risk of Bias</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>Overall Strength of Evidence</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preeclampsia</td>
<td>3 RCTs</td>
<td>Low</td>
<td>Consistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Moderate (favors treatment)</td>
<td>Difference in favor of treatment for RCTs (RR, 0.62 [95% CI, 0.43 to 0.89]); no difference observed for cohort study</td>
</tr>
<tr>
<td></td>
<td>1 cohort study</td>
<td>High</td>
<td>Unknown</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Maternal weight gain</td>
<td>4 RCTs</td>
<td>Medium</td>
<td>Inconsistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Insufficient</td>
<td>Results not pooled for RCTs because of substantial heterogeneity; no difference for cohort studies (MD, −1.04 [CI, −2.89 to 0.81])</td>
</tr>
<tr>
<td></td>
<td>2 cohort studies</td>
<td>High</td>
<td>Consistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Birth injury</td>
<td>2 RCTs</td>
<td>Medium</td>
<td>Consistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Low</td>
<td>Insufficient (favors treatment) No difference for RCTs (RR, 0.48 [CI, 0.12 to 1.90]); difference favoring treatment for cohort study (RR, 0.02 [CI, 0.00 to 0.22])</td>
</tr>
<tr>
<td></td>
<td>1 cohort study</td>
<td>High</td>
<td>Unknown</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia</td>
<td>3 RCTs</td>
<td>Medium</td>
<td>Consistent</td>
<td>Direct</td>
<td>Precise</td>
<td>Moderate (favors treatment)</td>
<td>No difference for RCTs (RR, 0.48 [CI, 0.12 to 1.90]); difference favoring treatment for cohort study (RR, 0.02 [CI, 0.00 to 0.22])</td>
</tr>
<tr>
<td></td>
<td>4 cohort studies</td>
<td>High</td>
<td>Consistent</td>
<td>Direct</td>
<td>Precise</td>
<td>Low (favors treatment)</td>
<td></td>
</tr>
<tr>
<td>Neonatal hypoglycemia</td>
<td>4 RCTs</td>
<td>Medium</td>
<td>Consistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Low (no difference) Insufficient</td>
<td>No difference for RCTs (RR, 1.18 [CI, 0.92 to 1.52]) or cohort studies (RR, 0.55 [CI, 0.10 to 2.97])</td>
</tr>
<tr>
<td></td>
<td>2 cohort studies</td>
<td>High</td>
<td>Inconsistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Macrosomia (birthweight &gt;4000 g)</td>
<td>5 RCTs</td>
<td>Medium</td>
<td>Consistent</td>
<td>Direct</td>
<td>Precise</td>
<td>Moderate (favors treatment)</td>
<td>No difference for RCTs (RR, 1.18 [CI, 0.92 to 1.52]) or cohort studies (RR, 0.55 [CI, 0.10 to 2.97])</td>
</tr>
<tr>
<td></td>
<td>6 cohort studies</td>
<td>High</td>
<td>Inconsistent</td>
<td>Direct</td>
<td>Precise</td>
<td>Low (favors treatment)</td>
<td></td>
</tr>
<tr>
<td>Long-term metabolic outcomes: impaired glucose tolerance</td>
<td>1 RCT</td>
<td>Medium</td>
<td>Unknown</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Insufficient</td>
<td>No difference between groups (RR, 5.63 [CI, 0.31 to 101.32])</td>
</tr>
<tr>
<td>Long-term metabolic outcomes: type 2 DM</td>
<td>1 RCT</td>
<td>Medium</td>
<td>Unknown</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Insufficient</td>
<td>No difference between groups (RR, 1.88 [CI, 0.08 to 44.76])</td>
</tr>
<tr>
<td>Long-term metabolic outcomes: BMI (assessed as &gt;85th and &gt;95th percentile)</td>
<td>2 RCTs</td>
<td>Medium</td>
<td>Consistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Low (no difference)</td>
<td>No difference between groups (RR, 1.26 [CI, 0.86 to 1.84])</td>
</tr>
</tbody>
</table>

BMI = body mass index; DM = diabetes mellitus; GDM = gestational diabetes mellitus; MD = mean difference; RCT = randomized, controlled trial; RR = risk ratio.
One RCT reported on BMI at delivery and showed lower BMI with treatment; however, this evidence was considered insufficient. There was no evidence from the included studies for long-term maternal outcomes, such as type 2 diabetes mellitus, obesity, and hypertension.

The Table and Appendix Table 3 (available at www.annals.org) show the findings for fetal, neonatal, or child outcomes. Evidence was insufficient for birth injury due to imprecision (low number of events and participants across studies) and inconsistency (2 RCTs showed no difference [12, 15], and 1 cohort study showed fewer cases with treatment [18]). Moderate evidence showed fewer cases of shoulder dystocia with treatment (Figure 1). For other injury outcomes (that is, brachial plexus injury and clavicular fractures), results were inconsistent across study designs, with the RCTs showing no differences and the cohort study showing fewer cases with treatment.

For outcomes related to birthweight (including birthweight >4000 g, actual birthweight, and large-for-gestational-age neonates), lower weights or fewer cases were observed with treatment. The strength of evidence was moderate for birthweight >4000 g (Figure 1). There was no difference in hyperbilirubinemia for RCTs (low strength of evidence), whereas the cohort study showed significantly less hyperbilirubinemia in the treated group.

There were no differences in perinatal death, although the number of events was extremely low (<0.5%). Randomized, controlled trials showed no difference between groups for the respiratory distress syndrome, whereas 1 cohort study found fewer “respiratory complications” (17) in

### Table 1. Effect of treatment for shoulder dystocia, neonatal hypoglycemia, and macrosomia (birthweight >4000 g) based on data from randomized, controlled trials.

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Events/Total, n/N</th>
<th>Weight, %</th>
<th>Risk Ratio MH, Random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shoulder dystocia</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bevier et al, 1999 (14)</td>
<td>1/35</td>
<td>6.4</td>
<td>0.69 (0.06–7.27)</td>
</tr>
<tr>
<td>Crowther et al, 2005 (6)</td>
<td>7/506</td>
<td>45.9</td>
<td>0.45 (0.19–1.09)</td>
</tr>
<tr>
<td>Landon et al, 2009 (12)</td>
<td>7/476</td>
<td>47.7</td>
<td>0.37 (0.16–0.88)</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>1017</td>
<td>100.0</td>
<td>0.42 (0.23–0.77)</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Heterogeneity: τ² = 0.00; χ² = 0.27; P = 0.87; I² = 0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 2.83 (P = 0.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Neonatal hypoglycemia</strong></th>
<th>Events/Total, n/N</th>
<th>Weight, %</th>
<th>Risk Ratio MH, Random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bevier et al, 1999 (14)</td>
<td>5/150</td>
<td>4.5</td>
<td>0.83 (0.26–2.67)</td>
</tr>
<tr>
<td>Crowther et al, 2005 (6)</td>
<td>35/506</td>
<td>25.9</td>
<td>1.34 (0.82–2.18)</td>
</tr>
<tr>
<td>Garner et al, 1997 (15)</td>
<td>21/149</td>
<td>14.4</td>
<td>1.63 (0.85–3.13)</td>
</tr>
<tr>
<td>Landon et al, 2009 (12)</td>
<td>62/381</td>
<td>55.3</td>
<td>1.06 (0.76–1.47)</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>1186</td>
<td>100.0</td>
<td>1.18 (0.92–1.52)</td>
</tr>
<tr>
<td>Total</td>
<td>123</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>Heterogeneity: τ² = 0.00; χ² = 1.96; P = 0.58; I² = 0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 1.33 (P = 0.18)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Macrosomia (&gt;4000 g)</strong></th>
<th>Events/Total, n/N</th>
<th>Weight, %</th>
<th>Risk Ratio MH, Random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bevier et al, 1999 (14)</td>
<td>1/35</td>
<td>2.9</td>
<td>0.11 (0.02–0.84)</td>
</tr>
<tr>
<td>Bonomo et al, 2005 (13)</td>
<td>8/150</td>
<td>13.1</td>
<td>0.50 (0.22–1.13)</td>
</tr>
<tr>
<td>Crowther et al, 2005 (6)</td>
<td>49/506</td>
<td>33.1</td>
<td>0.46 (0.34–0.63)</td>
</tr>
<tr>
<td>Garner et al, 1997 (15)</td>
<td>24/149</td>
<td>23.7</td>
<td>0.86 (0.53–1.42)</td>
</tr>
<tr>
<td>Landon et al, 2009 (12)</td>
<td>28/477</td>
<td>27.2</td>
<td>0.41 (0.27–0.63)</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>1317</td>
<td>100.0</td>
<td>0.50 (0.35–0.71)</td>
</tr>
<tr>
<td>Total</td>
<td>110</td>
<td>231</td>
<td></td>
</tr>
<tr>
<td>Heterogeneity: τ² = 0.07; χ² = 7.94; P = 0.09; I² = 50%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 3.84 (P &lt; 0.001)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MH = Mantel–Haenszel.
the treated group; overall respiratory distress syndrome was rare (4.3% across all studies). Several studies assessed Apgar scores; although differences were found for the Apgar score at 1 minute, no differences were observed at 5 minutes.

One RCT followed a subset of the offspring for 7 to 11 years and found no differences for impaired glucose tolerance or type 2 diabetes mellitus (insufficient strength of evidence). No differences were observed in single studies that assessed offspring with BMIs greater than the 95th percentile (7- to 11-year follow-up) and greater than the 85th percentile (4- to 5-year follow-up). Overall, pooled results showed no difference in BMI (low strength of evidence).

**Harms of Treating GDM**

One RCT assessed maternal depression and anxiety at 6 weeks after study entry and 3 months after parturition (6). There was no difference between groups in anxiety at either time point. Depression rates were lower in the treatment group 3 months after parturition (Appendix Table 2).

Moderate evidence from 4 RCTs showed no difference in small-for-gestational-age neonates. Pooled results from 4 RCTs showed no difference between groups in neonatal hypoglycemia and no statistical heterogeneity (Figure 1). Two cohort studies showed inconsistent results, which may be partly due to different definitions of hypoglycemia used across the studies and different protocols for screening neonates for hypoglycemia. Overall, the strength of evidence was low, suggesting that further study may change the results of our findings (Table).

Low evidence showed no difference overall in admission to the neonatal intensive care unit (Appendix Figure 4, available at www.annals.org). One trial was an outlier, with significantly more neonatal intensive care unit admissions in the treated group. Two RCTs reported on the number of prenatal visits and found more visits among the treatment groups. The strength of evidence for induction of labor was insufficient because of lack of precision and inconsistency across studies, with no difference found for the RCTs overall. There was low evidence of no differences between groups for cesarean delivery (Figure 2) or unplanned cesarean delivery.

**DISCUSSION**

Moderate evidence showed that treatment of GDM reduced preeclampsia, shoulder dystocia, and macrosomia (birthweight >4000 g). These outcomes, specified a priori

---

**Figure 2. Effect of treatment on outcomes of women with GDM who have cesarean delivery.**

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Events/Total, n/N</th>
<th>Weight, %</th>
<th>Risk Ratio MH, Random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RCTs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bevier et al, 1999 (14)</td>
<td>5/35</td>
<td>1.6</td>
<td>0.57 (0.22–1.47)</td>
</tr>
<tr>
<td>Bonomo et al, 2005 (13)</td>
<td>42/150</td>
<td>11.3</td>
<td>0.95 (0.67–1.36)</td>
</tr>
<tr>
<td>Crowther et al, 2005 (6)</td>
<td>152/490</td>
<td>43.1</td>
<td>0.96 (0.80–1.16)</td>
</tr>
<tr>
<td>Garner et al, 1997 (15)</td>
<td>30/149</td>
<td>6.7</td>
<td>1.08 (0.68–1.71)</td>
</tr>
<tr>
<td>Landon et al, 2009 (12)</td>
<td>128/476</td>
<td>37.3</td>
<td>0.79 (0.65–0.97)</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>1300</td>
<td>100.0</td>
<td>0.90 (0.79–1.01)</td>
</tr>
<tr>
<td>Total</td>
<td>357</td>
<td>402</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: tau-square = 0.00; chi-square = 3.68; P = 0.45; I² = 0%

Test for overall effect: Z = 1.81 (P = 0.07)

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Events/Total, n/N</th>
<th>Weight, %</th>
<th>Risk Ratio MH, Random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cohort studies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adams et al, 1998 (18)</td>
<td>99/373</td>
<td>4.5</td>
<td>1.06 (0.45–2.52)</td>
</tr>
<tr>
<td>Bonomo et al, 1997 (19)</td>
<td>7/26</td>
<td>6.4</td>
<td>0.91 (0.45–1.85)</td>
</tr>
<tr>
<td>Chou et al, 2010 (21)</td>
<td>40/233</td>
<td>15.0</td>
<td>1.74 (1.13–2.69)</td>
</tr>
<tr>
<td>Fassett et al, 2007 (16)</td>
<td>21/69</td>
<td>11.5</td>
<td>0.91 (0.55–1.52)</td>
</tr>
<tr>
<td>Langer et al, 2005 (17)</td>
<td>258/1110</td>
<td>43.0</td>
<td>0.98 (0.81–1.17)</td>
</tr>
<tr>
<td>Naylor et al, 1996 (20)</td>
<td>48/143</td>
<td>19.6</td>
<td>1.14 (0.79–1.63)</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>1954</td>
<td>100.0</td>
<td>1.09 (0.90–1.31)</td>
</tr>
<tr>
<td>Total</td>
<td>473</td>
<td>247</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: tau-square = 0.01; chi-square = 6.47; P = 0.26; I² = 23%

Test for overall effect: Z = 0.88 (P = 0.38)

Test for subgroup differences: chi-square = 2.93; P = 0.09; I² = 65.9%

GDM = gestational diabetes mellitus; MH = Mantel–Haenszel; RCT = randomized, controlled trial.
to be of interest to our stakeholders, may be intermediate
to outcomes of greater clinical importance, such as prematurity
or brachial plexus injury. Evidence showing differences
between groups for other benefits to mother or infant
was lacking or weak.

In terms of harms, there was no evidence for some of
the outcomes stipulated in the protocol, including costs
and resource allocation, although there were more prenatal
visits in the treatment groups. No difference was found in
small-for-gestational-age neonates, which may be due to
inadequate power to detect differences because of the small
number of events. No differences were found for admission
to the neonatal intensive care unit or rate of induction of
labor. However, there was heterogeneity in these outcomes
that may be attributable to different site-specific policies
and procedures, study protocols, and practice patterns.
Low evidence showed no difference in rates of cesarean
delivery.

Our results are consistent with other recent systematic
reviews showing some evidence of benefit of treating GDM
for select maternal and infant outcomes yet little evidence
of an effect on patient-important outcomes (for example,
perinatal or neonatal mortality). This is probably due to
the infrequent occurrence of these events and a resulting
lack of power across the studies to adequately assess for
differences (24, 25).

Several caveats related to this body of evidence should
be considered when interpreting and applying the results
of this review. First, although we found differences in pre-
eclampsia, macrosomia, and shoulder dystocia, most such
events occur in pregnant women without GDM (26). Such
factors as maternal weight and gestational weight gain have
been shown to impart greater risk for these outcomes, par-
ticularly in women diagnosed with GDM at lower glucose
thresholds (27, 28). For example, analyses adjusting for
these variables show that glycaemia accounted for only
1.7% of the risk for large-for-gestational-age neonates (27).
Second, where reported, definitions of preeclampsia varied
(for example, a blood pressure of 140/90 mm Hg on 2
occasions 4 hours apart, these criteria with laboratory mea-
sures indicative of preeclampsia, or an increase in blood
pressure medications). Preeclampsia events in our pooled
analysis may have included women with the much-less-
serious condition of gestational hypertension; however, a
study that used the more rigorous definition showed a
treatment benefit (12). Preeclampsia occurs in 3% to 5%
of pregnancies (29), and the risk for this condition attribu-
table to GDM is probably small (30). Third, this review
assessed the risks and benefits of treating GDM but not
those of screening for this condition. Of note, our larger
technical report, which addressed screening, found no ran-
domized trials examining the effect of screening on health
outcomes.

Evidence was very limited for 2 outcomes of particular
interest to stakeholders. The first was patient anxiety asso-
ciated with a diagnosis of GDM. A single study assessed
depression and anxiety in a subgroup of a larger RCT. It
found no difference between groups in anxiety at 6 weeks
after study entry and 3 months after parturition, although
the treatment group had lower rates of depression at 3
months after parturition. Research has shown that women
with GDM had a higher level of anxiety at the time of the
first GDM assessment than glucose-tolerant women; how-
ever, these differences in anxiety scores did not persist be-
fore delivery (31). Further, a survey of women 3 to 5 years
after diagnosis of GDM showed more concern about their
own health and rated their children’s health poorer than
matched control participants (32). The second outcome
was metabolic changes in the children born to mothers
with GDM. Follow-up of offspring from participants in 2
RCTs (6, 16) did not show any treatment effect of GDM
on metabolic outcomes of the children.

Further study of the long-term metabolic effect on
offspring whose mothers have been treated for GDM is
warranted. Well-conducted prospective cohort studies of
the real-world effect of GDM treatment on health care
utilization are needed. Research is also needed to help de-
termine the glucose thresholds and treatment targets at
which GDM treatment benefits outweigh the risks of treat-
ment and no treatment.

The IDEAL (Investigation of Dietary Advice and Life-
style for Women With Borderline Gestational Diabetes)
study, an RCT to assess the effect of treating women with
very mild glucose impairment in pregnancy, is under way.
Randomized, controlled trials investigating the care of
women diagnosed with GDM, including fetal surveillance
protocols, are needed to guide obstetric investigations and
management of GDM. Such work may help avoid unnec-
essary interventions that are driven by the apprehension
of health care providers.

The review process had several limitations. We limited
the search dates from 1995 onward on the basis of advice
from our technical expert panel. Our results are consistent
with other systematic reviews on this topic that included
studies before 1995 (24, 25). We included only studies
published in English. Most studies were conducted in
North America or Australia. Most of the North American
studies included mixed racial populations and are probably
applicable to the general U.S. population. We included
cohort studies because an earlier review (5) found few
RCTs; results from cohort studies should be interpreted
cautiously, particularly when they differ from those of the
RCTs.

In summary, evidence supports benefits of treating
mild GDM. Specifically, treatment of GDM results in
lower incidence of preeclampsia, macrosomia, large-for-
gestational-age infants, and shoulder dystocia; however, the
risk for these outcomes attributable to GDM is low, par-
ticularly when glucose levels are modestly elevated. Current
research does not show a treatment effect of GDM on clinical neonatal hypoglycemia or future poor metabolic outcomes of the offspring. Randomized, controlled trials of GDM treatment show limited harm related to treating GDM, other than an increased demand for services.

From the Alberta Research Centre for Health Evidence and the University of Alberta Evidence-based Practice Center, University of Alberta, Edmonton, Alberta, Canada, and the University of Calgary, Calgary, Alberta, Canada.

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

Acknowledgment: The authors thank Tamara Durec (searching), Andrea Milne (searching, technical support), Walie Atkary (screening, data extraction), Dion Pasichnyk (data extraction), Jennifer Seida (screening, project coordination), Teodora Radiс (article retrieval), Jocelyn Shulhan (screening), Annabritt Chisholm (research support, reference management), and Dr. Alun Edwards for providing clinical input throughout the project.

Financial Support: By the AHRQ (contract no. 290-2007-10021-I).

Potential Conflicts of Interest: Ms. Guthrie: Grant (money to institution); support for travel to meetings for the study or other purposes (money to institution); fees for participation in review activities such as data monitoring boards, statistical analysis, end point committees, and the like (money to institution); payment for writing or reviewing the manuscript (money to institution); AHRQ, Ms. Muse: Grant (money to institution); AHRQ, Mr. Vandermeer: Grant (money to institution); AHRQ, Dr. Donovan: Other: Dr. Donovan reports grants from Bridges Grant (International Diabetes Federation and Eli Lilly) outside the submitted work. All other authors have no disclosures. Disclosures can also be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M13-0190.

Requests for Single Reprints: Lisa Hartling, PhD, ECHA 4-472, 11405-87 Avenue, Edmonton, AB T6G 1C9, Canada; e-mail, hartling@ualberta.ca.

Current author addresses and author contributions are available at www.annals.org.

References
27. Black MH, Sacks DA, Xiang AH, Lawrence JM. The relative contribution of prepregnancy overweight and obesity, gestational weight gain, and IADPSG-


**Current Author Addresses:** Dr. Hartling: ECHA 4-472, 11405-87 Avenue, Edmonton, AB T6G 1C9, Canada.
Dr. Dryden: ECHA 4-474, 11405-87 Avenue, Edmonton, AB T6G 1C9, Canada.
Ms. Guthrie: Primary Care Division, Alberta Health, 18th Floor, Telus Plaza North Tower, 10025 Jasper Avenue NW, Edmonton, AB T5J 1S6, Canada.
Ms. Muise: ECHA 4-492C, 11405-87 Avenue, Edmonton, AB T6G 1C9, Canada.
Mr. Vandermeer: ECHA 4-496B, 11405-87 Avenue, Edmonton, AB T6G 1C9, Canada.
Dr. Donovan: Richmond Road Diagnostic and Treatment Centre, 1820 Richmond Road SW, Calgary, AB T2T 5C7, Canada.

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

1. Diabetes, Gestational/
2. Fetal Macrosomia/
3. Pregnancy Complications/
4. GDM.tw.
5. (gestation$ adj2 (diabet$ or DM or glucose intoleran$ or insulin resistan$)).mp.
6. (pregnan$ adj3 (diabet$ or DM or glucose intoleran$ or insulin resistan$)).mp.
7. (maternal adj2 (diabet$ or DM or gly?emia or hyperglyc?emia)).tw.
9. macrosomia.tw.
10. or/1-9
11. mass screening/
12. prenatal diagnosis/
13. screen$.tw.
14. ((prenatal or early) adj2 diagnosis).tw.
15. Glucose Tolerance Test/
16. Glucose Intolerance/
17. Blood Glucose/
18. Risk Factors/
19. (glucose adj (tolerance or intolerance or challenge)).tw.
20. OGGT.tw.
21. GCT.tw.
22. (fasting adj2 glucose).tw.
23. or/11-22
24. “Sensitivity and Specificity” /
25. “Predictive Value of Tests” /
26. ROC Curve/
27. specificity.tw.
28. sensitivity.tw.
29. predictive value.tw.
30. accuracy.tw.
31. diagnostic errors/
32. diagnostic error?.tw.
33. false negative reactions/
34. false positive reactions/
35. (false adj (negative or positive)).tw.
36. “reproducibility of results” /
37. reference values/
38. reference standards/
39. or/24-38
40. and/10,23,39
41. intervention?.mp.
42. (treatment or treatment? or therapy or therapies).mp.
43. manage$.mp.
44. monitor$.mp.
45. exp sulfonylurea compounds/
46. Gliclazide/
47. Glyburide/
48. Tolbutamide/
49. sulfonylurea?.tw.
50. glinide$tw.
51. glibizide$.tw.
52. glyburid$.tw.
53. tobutamid$.tw.
54. (antidiabet$ or anti-diabet$).tw.
55. insulin?.mp.
56. glibenclamid$.mp.
57. acarbose.mp.
58. exp Diet Therapy/
59. (diet adj2 (therap$ or restrict$ or advice)).tw.
60. medical nutrition$ therapy.t.
61. MNT.tw.
62. exp Life Style/
63. (lifestyle$ or life-style$).mp.
64. Blood Glucose Self-Monitoring/
65. (blood glucose adj (self monitor$ or self-monitor$)).tw.
66. ((self monitor$ or self-monitor$) adj blood glucose).tw.
67. SMBG.tw.
68. Counseling/

Appendix Table 1—Continued

70. counsel$.tw.
71. Labor, Induced/
73. exp Cesarean Section/
74. cesarean.tw.
75. exp Pregnancy Outcome/
76. pregnanc$ outcome?.tw.
77. or/41-76
78. and/10,77
79. or/40,78
80. clinical trial.pt.
81. randomized controlled trial.pt.
82. random?.ed.ti,ab.
83. placebo.ti,ab.
84. dt.fs.
85. randomly ti,ab.
86. trial.ti,ab.
87. groups.ti,ab.
88. or/80-87
89. animals/
90. humans/
91. 89 not (89 and 90)
92. 88 not 91
93. cohort studies/
94. follow-up studies/
95. longitudinal studies/
96. prospective studies/
97. retrospective studies/
98. ((cohort? or follow-up or followup or longitud$ or prospectiv$ or retrospectiv$) adj (study or studies or trial?)).tw.
99. or/93-98
100. 99 not 91
101. exp Guideline/
102. Health Planning Guidelines/
103. (clinical adj2 guideline$).tw.
104. CPG?.tw.
105. ((practice or consensus or position) adj2 (guideline? or recommendation? or statement?)).tw.
106. standard?.tw.
107. protocol?.tw.
108. or/101-107
109. meta analysis.mp.pt.
110. review.pt.
111. search:.tw.
112. or/109-111 [Reviews balanced - HIRU]
113. and/79,92 [Clinical trials & RCTs]
114. and/79,100 [Observational studies]
115. and/79,108 [Guidelines]
116. and/79,112 [SRs MAs]
117. or/113-116
118. limit 117 to (english language and yr="2000 -Current")
119. limit 117 to (english language and yr="2000 -2005")
120. limit 117 to (english language and yr="2006 -Current")
121. remove duplicates from 119
122. remove duplicates from 120
123. or/121-122
124. 113 or 114 or 115
125. 113 or 114 or 115
126. limit 125 to (english language and yr="2000 -Current")
127. limit 125 to (english language and yr="2000 -2005")
128. remove duplicates from 127
129. limit 125 to (english language and yr="2006 -Current")
130. remove duplicates from 129
131. 128 or 130
132. 113 or 114
133. limit 132 to (english language and yr="2000 -Current")
134. limit 132 to (english language and yr="2000 -2005")
135. remove duplicates from 134
136. limit 132 to (english language and yr="2006 -Current")
137. remove duplicates from 136
138. 135 or 137

* Database. Medline (Ovid interface): 1948 to week 4 September 2011; search date: 9 October 2011; results, 8234.
Appendix Figure 1. Summary of evidence search and selection.

- Total number of citations retrieved from electronic literature searches \( (n = 14,398) \)
- References selected for further examination of titles and abstracts \( (n = 598) \)
  - Potentially relevant references identified by hand-searching \( (n = 30) \)
  - Not retrieved \( (n = 8) \)
- Articles retrieved and evaluated for inclusion \( (n = 620) \)
  - Included \( (n = 151) \)
  - Excluded \( (n = 469) \)
    - Multiple publications \( (n = 26) \)
  - Unique studies \( (n = 125) \)
  - Excluded during extraction
    - No comparison or outcome of interest found during extraction \( (n = 28) \)
- Extracted studies \( (n = 97) \)
  - Studies addressing objectives of this review \( (n = 11) \)

KQ = key question.* This systematic review was part of a larger technical report. The search was done to identify relevant studies for all objectives of the full report, which is available at [http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=H11005](http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=H11005)
## Appendix Table 2. Evidence Summary for Benefits and Harms of Treating GDM: Maternal Outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Studies, n</th>
<th>Participants, n</th>
<th>Effect Estimate Risk Ratio (95% CI)*</th>
<th>I², %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benefits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preeclampsia</td>
<td>RCT</td>
<td>3 2014</td>
<td>0.62 (0.43 to 0.89)†</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Cohort</td>
<td>1 258</td>
<td>0.97 (0.43 to 2.15)</td>
<td>NA</td>
</tr>
<tr>
<td>Preeclampsia or gestational hypertension</td>
<td>RCT</td>
<td>1 931</td>
<td>0.63 (0.44 to 0.92)†</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Cohort</td>
<td>1 874</td>
<td>0.30 (0.15 to 0.62)†</td>
<td>NA</td>
</tr>
<tr>
<td>Weight gain (kg)</td>
<td>RCT</td>
<td>4 2530</td>
<td>Pooled estimate not reported because of heterogeneity</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Cohort</td>
<td>2 515</td>
<td>−1.04 (−2.89 to 0.81)‡</td>
<td>8</td>
</tr>
<tr>
<td>Maternal birth trauma</td>
<td>Cohort</td>
<td>1 874</td>
<td>0.95 (0.21 to 4.28)</td>
<td>NA</td>
</tr>
<tr>
<td>BMI at delivery</td>
<td>RCT</td>
<td>1 931</td>
<td>−1.00 (−1.67 to −0.33)†‡</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Harms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cesarean delivery</td>
<td>RCT</td>
<td>5 2613</td>
<td>0.90 (0.79 to 1.01)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Cohort</td>
<td>6 3110</td>
<td>1.09 (0.90 to 1.31)</td>
<td>23</td>
</tr>
<tr>
<td>Unplanned cesarean delivery</td>
<td>RCT</td>
<td>1 1000</td>
<td>0.81 (0.62 to 1.05)</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Cohort</td>
<td>1 126</td>
<td>0.83 (0.33 to 2.06)</td>
<td>NA</td>
</tr>
<tr>
<td>Induction of labor</td>
<td>RCT</td>
<td>2 1931</td>
<td>1.16 (0.91 to 1.49)</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Cohort</td>
<td>1 1665</td>
<td>0.63 (0.55 to 0.72)§</td>
<td>NA</td>
</tr>
<tr>
<td>Anxiety (6 wk after study entry)</td>
<td>RCT</td>
<td>1 682</td>
<td>−0.30 (−0.88 to 0.28)</td>
<td>NA</td>
</tr>
<tr>
<td>Anxiety (3 mo after parturition)</td>
<td>RCT</td>
<td>1 573</td>
<td>−0.20 (−0.83 to 0.43)</td>
<td>NA</td>
</tr>
<tr>
<td>Depression (3 mo after parturition)</td>
<td>RCT</td>
<td>1 568</td>
<td>0.50 (0.31 to 0.79)†</td>
<td>NA</td>
</tr>
</tbody>
</table>

BMI = body mass index; GDM = gestational diabetes mellitus; NA = not applicable; RCT = randomized, controlled trial.

* Risk ratios unless otherwise specified.
† Statistically significant with better results for the treated group.
‡ Mean difference.
§ This result was statistically significant; however, all untreated women in this cohort presented at or after 37 wks’ gestation, and institutional policy required that such women be delivered within 1 wk of presentation.
Appendix Figure 2. Effect of treatment on outcomes of women with GDM: preeclampsia.

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Events/Total, n/N</th>
<th>Weight, %</th>
<th>Risk Ratio MH, Random (95% CI)</th>
<th>Risk Ratio MH, Random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td>No Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCTs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bevier et al, 1999 (14)</td>
<td>2/35</td>
<td>1/48</td>
<td>1.8</td>
<td>2.74 (0.26–29.07)</td>
</tr>
<tr>
<td>Crowther et al, 2005 (6)</td>
<td>58/490</td>
<td>93/510</td>
<td>65.1</td>
<td>0.65 (0.48–0.88)</td>
</tr>
<tr>
<td>Landon et al, 2009 (12)</td>
<td>12/476</td>
<td>25/455</td>
<td>19.1</td>
<td>0.46 (0.23–0.90)</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>100/101</td>
<td>1013</td>
<td>85.9</td>
<td>0.62 (0.43–0.89)</td>
</tr>
<tr>
<td>Total</td>
<td>72</td>
<td>119</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: tau-square = 0.02; chi-square = 2.37; P = 0.31; I² = 16%
Test for overall effect: Z = 2.57 (P = 0.01)

Cohort studies

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Events/Total, n/N</th>
<th>Weight, %</th>
<th>Risk Ratio MH, Random (95% CI)</th>
<th>Risk Ratio MH, Random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naylor et al, 1996 (20)</td>
<td>12/143</td>
<td>10/115</td>
<td>14.1</td>
<td>0.97 (0.43–2.15)</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>143</td>
<td>115</td>
<td>14.1</td>
<td>0.97 (0.43–2.15)</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: Not applicable
Test for overall effect: Z = 0.09 (P = 0.93)

Total (95% CI) 1144 1128 100.0 0.66 (0.48–0.90)
Total 84 129

Heterogeneity: tau-square = 0.02; chi-square = 3.38; P = 0.34; I² = 11%
Test for overall effect: Z = 2.60 (P = 0.009)
Test for subgroup differences: chi-square = 0.99; P = 0.32; I² = 0%

GDM = gestational diabetes mellitus; MH = Mantel–Haenszel; RCT = randomized, controlled trial.

Appendix Figure 3. Effect of treatment on outcomes of women with GDM: maternal weight gain.

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Treatment Mean (SD)</th>
<th>Total</th>
<th>No Treatment Mean (SD)</th>
<th>Total</th>
<th>Mean Difference IV, Random (95% CI)</th>
<th>Mean Difference IV, Random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td></td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCTs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonomo et al, 2005 (13)</td>
<td>13.1 (4.3)</td>
<td>150</td>
<td>12.6 (3.9)</td>
<td>150</td>
<td>0.50 (–0.43 to 1.43)</td>
<td></td>
</tr>
<tr>
<td>Crowther et al, 2005 (6)</td>
<td>8.1 (0.3)</td>
<td>490</td>
<td>9.8 (0.4)</td>
<td>510</td>
<td>–1.70 (–1.74 to –1.66)</td>
<td></td>
</tr>
<tr>
<td>Garner et al, 1997 (15)</td>
<td>12.54 (16.50)</td>
<td>149</td>
<td>13.37 (19.90)</td>
<td>150</td>
<td>–0.83 (–4.97 to 3.31)</td>
<td></td>
</tr>
<tr>
<td>Landon et al, 2009 (12)</td>
<td>2.8 (4.5)</td>
<td>476</td>
<td>5.0 (3.3)</td>
<td>455</td>
<td>–2.20 (–2.71 to –1.69)</td>
<td></td>
</tr>
<tr>
<td>Cohort studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adams et al, 1998 (18)</td>
<td>12.26 (7.09)</td>
<td>373</td>
<td>14.24 (4.90)</td>
<td>16</td>
<td>–1.98 (–4.49 to 0.53)</td>
<td></td>
</tr>
<tr>
<td>Fassett et al, 2007 (16)</td>
<td>10.34 (8.8)</td>
<td>69</td>
<td>10.43 (5.49)</td>
<td>57</td>
<td>–0.09 (–2.61 to 2.43)</td>
<td></td>
</tr>
</tbody>
</table>

GDM = gestational diabetes mellitus; IV = inverse variance; MH = Mantel–Haenszel; RCT = randomized, controlled trial.
## Appendix Table 3. Evidence Summary for Benefits and Harms of Treating GDM: Infant Outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Studies, n</th>
<th>Participants, n</th>
<th>Effect Estimate Risk Ratio (95% CI)*</th>
<th>I², %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benefits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birthweight &gt;4000 g</td>
<td>5</td>
<td>2643</td>
<td>0.50 (0.35 to 0.71)†</td>
<td>50</td>
</tr>
<tr>
<td>Cohort</td>
<td>6</td>
<td>3426</td>
<td>Results not pooled because of substantial heterogeneity</td>
<td>86</td>
</tr>
<tr>
<td>Birthweight &gt;4500 g</td>
<td>1</td>
<td>299</td>
<td>1.01 (0.33 to 3.05)</td>
<td>NA</td>
</tr>
<tr>
<td>Cohort</td>
<td>2</td>
<td>647</td>
<td>0.29 (0.07 to 1.25)</td>
<td>69</td>
</tr>
<tr>
<td>Birthweight (actual)</td>
<td>5</td>
<td>2670</td>
<td>−120.81 (−163.40 to −78.23)‡</td>
<td>2</td>
</tr>
<tr>
<td>Cohort</td>
<td>2</td>
<td>515</td>
<td>Results not pooled because of substantial heterogeneity</td>
<td>77</td>
</tr>
<tr>
<td>Large-for-gestational-age neonate</td>
<td>3</td>
<td>2261</td>
<td>0.56 (0.45 to 0.69)†</td>
<td>0</td>
</tr>
<tr>
<td>Cohort</td>
<td>4</td>
<td>2294</td>
<td>0.43 (0.27 to 0.70)†</td>
<td>58</td>
</tr>
<tr>
<td>Shoulder dystocia</td>
<td>3</td>
<td>2044</td>
<td>0.42 (0.23 to 0.77)†</td>
<td>0</td>
</tr>
<tr>
<td>Brachial plexus injury</td>
<td>4</td>
<td>3054</td>
<td>0.38 (0.19 to 0.78)†</td>
<td>20</td>
</tr>
<tr>
<td>Clavicular fracture</td>
<td>1</td>
<td>1000</td>
<td>0.15 (0.01 to 2.87)</td>
<td>NA</td>
</tr>
<tr>
<td>Cohort</td>
<td>1</td>
<td>389</td>
<td>0.04 (0.00 to 0.66)†</td>
<td>NA</td>
</tr>
<tr>
<td>Respiratory complications</td>
<td>3</td>
<td>2287</td>
<td>−0.00 (−0.01 to 0.01)§</td>
<td>66</td>
</tr>
<tr>
<td>Cohort</td>
<td>3</td>
<td>2928</td>
<td>−0.00 (−0.01 to 0.01)§</td>
<td>0</td>
</tr>
<tr>
<td>Respiratory complications (RDS)</td>
<td>2</td>
<td>1962</td>
<td>1.05 (0.48 to 2.28)</td>
<td>58</td>
</tr>
<tr>
<td>Cohort (complications)</td>
<td>1</td>
<td>1665</td>
<td>0.16 (0.10 to 0.26)†</td>
<td>NA</td>
</tr>
<tr>
<td>Apgar score at 1 min</td>
<td>1</td>
<td>83</td>
<td>−0.30 (−0.56 to −0.04)‡</td>
<td>NA</td>
</tr>
<tr>
<td>Cohort</td>
<td>1</td>
<td>126</td>
<td>−1.00 (−1.54 to −0.46)‡</td>
<td>NA</td>
</tr>
<tr>
<td>Apgar score at 5 min</td>
<td>2</td>
<td>383</td>
<td>Results not pooled because of substantial heterogeneity</td>
<td>77</td>
</tr>
<tr>
<td>Cohort</td>
<td>1</td>
<td>126</td>
<td>0.00 (−0.27 to 0.27)‡</td>
<td>NA</td>
</tr>
<tr>
<td>Type 2 DM (long-term)</td>
<td>1</td>
<td>89</td>
<td>1.88 (0.08 to 44.76)</td>
<td>NA</td>
</tr>
<tr>
<td>Impaired glucose tolerance (RCT)</td>
<td>1</td>
<td>89</td>
<td>5.63 (0.31 to 101.32)</td>
<td>44</td>
</tr>
<tr>
<td>BMI (long-term)</td>
<td>1</td>
<td>85</td>
<td>1.58 (0.66 to 3.79)</td>
<td>NA</td>
</tr>
<tr>
<td>&gt;95th percentile</td>
<td>1</td>
<td>199</td>
<td>1.19 (0.78 to 1.82)</td>
<td>NA</td>
</tr>
<tr>
<td>&gt;85th percentile</td>
<td>1</td>
<td>199</td>
<td>1.19 (0.78 to 1.82)</td>
<td>NA</td>
</tr>
<tr>
<td>Any BMI (2 studies above combined)</td>
<td>2</td>
<td>284</td>
<td>1.26 (0.86 to 1.84)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Harms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-for-gestational-age neonate</td>
<td>4</td>
<td>2345</td>
<td>1.10 (0.81 to 1.48)</td>
<td>0</td>
</tr>
<tr>
<td>Hypoglycemia</td>
<td>4</td>
<td>2367</td>
<td>1.18 (0.92 to 1.52)</td>
<td>0</td>
</tr>
<tr>
<td>Admission to NICU</td>
<td>2</td>
<td>2054</td>
<td>0.55 (0.10 to 2.97)</td>
<td>49</td>
</tr>
<tr>
<td>RCT</td>
<td>3</td>
<td>2262</td>
<td>0.96 (0.67 to 1.37)</td>
<td>61</td>
</tr>
<tr>
<td>Cohort</td>
<td>1</td>
<td>126</td>
<td>0.66 (0.19 to 2.35)</td>
<td>NA</td>
</tr>
</tbody>
</table>

BMI = body mass index; DM = diabetes mellitus; GDM = gestational diabetes mellitus; NA = not applicable; NICU = neonatal intensive care unit; RCT = randomized, controlled trial; RDS = respiratory distress syndrome.

* Risk ratios unless otherwise specified.
† Results statistically significant with more benefits for the treated group.
‡ Mean difference.
§ Risk difference.
Appendix Figure 4. Effect of treatment on outcomes of women with GDM: admission to the NICU.

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Treatment Events/Total, n/N (%)</th>
<th>Weight, %</th>
<th>Risk Ratio MH, Random (95% CI)</th>
<th>Risk Ratio MH, Random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCTs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonomo et al, 2005 (13)</td>
<td>5/150</td>
<td>8.6</td>
<td>0.71 (0.23–2.20)</td>
<td></td>
</tr>
<tr>
<td>Crowther et al, 2005 (6)</td>
<td>357/506</td>
<td>56.4</td>
<td>1.15 (1.05–1.26)</td>
<td></td>
</tr>
<tr>
<td>Landon et al, 2009 (12)</td>
<td>43/477</td>
<td>35.0</td>
<td>0.77 (0.53–1.13)</td>
<td></td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>1133</td>
<td>100.0</td>
<td>0.96 (0.67–1.37)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>405</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: tau-square = 0.06, chi-square = 5.16; P = 0.08; I² = 61%
Test for overall effect: Z = 0.21 (P = 0.83)

| Cohort studies          |                                 |           |                                |                                |
|-------------------------|                                 |           |                                |                                |
| Fassett et al, 2007 (16) | 4/69                            | 100.0     | 0.66 (0.19–2.35)               |                                |
| Subtotal (95% CI)       | 69                              | 100.0     | 0.66 (0.19–2.35)               |                                |
| Total                   | 4                               |           |                                |                                |

Heterogeneity: Not applicable
Test for overall effect: Z = 0.64 (P = 0.52)
Test for subgroup differences: chi-square = 0.31; P = 0.58; I² = 0%

GDM = gestational diabetes mellitus; MH = Mantel–Haenszel; NICU = neonatal intensive care unit; RCT = randomized, controlled trial.