Evidence Report/Technology Assessment

Number 53

Management of Prolonged Pregnancy

Prepared for:
Agency for Healthcare Research and Quality
2101 East Jefferson Street
Rockville, MD 20852
http://www.ahrq.gov

Contract No. 290-97-0014

Prepared by:
Duke Evidence-based Practice Center
Durham, NC

Evan R. Myers, MD, MPH
Richard Blumrick, MD
Andrea L. Christian, BA
Santanu Datta, MBA, MS
Rebecca N. Gray, DPhil
Jane T. Kolimaga, MA
Elizabeth Livingston, MD
Andrea Lukes, MD, MHSc
David B. Matchar, MD
Douglas C. McCrory, MD, MHSc

AHRQ Publication No. 02-E018
May 2002

Suggested citation:
Myers ER, Blumrick R, Christian AL, et al. Management of prolonged pregnancy. Evidence Report/Technology Assessment No. 53 (Prepared by Duke Evidence-based Practice Center, Durham, NC, under Contract No. 290-97-0014). AHRQ Publication No. 02-E018. Rockville, MD: Agency for Healthcare Research and Quality. May 2002.

On December 6, 1999, under Public Law 106-129, the Agency for Health Care Policy and Research (AHCPR) was reauthorized and renamed the Agency for Healthcare Research and Quality (AHRQ). The law authorizes AHRQ to continue its research on the cost, quality, and outcomes of health care, and expands its role to improve patient safety and address medical errors.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

Preface

The Agency for Healthcare Research and Quality (AHRQ, formerly the Agency for Health Care Policy and Research, AHCPR), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To bring the broadest range of experts into the development of evidence reports and health technology assessments, AHRQ encourages the EPCs to form partnerships and enter into collaborations with other medical and research organizations. The EPCs work with these partner organizations to ensure that the evidence reports and technology assessments they produce will become building blocks for health care quality improvement projects throughout the Nation. The reports undergo peer review prior to their release.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality.

We welcome written comments on this evidence report. They may be sent to: Director, Center for Practice and Technology Assessment, Agency for Healthcare Research and Quality, 6010 Executive Blvd., Suite 300, Rockville, MD 20852.

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

Robert Graham, M.D.
Director,
Center for Practice
and Technology Assessment

Structured Abstract

Objective. Approximately 18 percent of pregnancies in the United States extend beyond 41 weeks gestation, 7 percent beyond 42 weeks. Risks of adverse perinatal and maternal outcomes increase with increasing gestational age beyond term. This report assesses the literature on the benefits, risks, and costs of different strategies for managing prolonged pregnancy in order to avoid adverse perinatal and maternal outcomes.

Search Strategy. Published literature on the management of prolonged pregnancy was identified in MEDLINE, CINAHL, EMBASE, HealthSTAR, the Cochrane Database of Systematic Reviews, and the Database of Abstracts of Reviews of Effectiveness for the years 1980 through 2001. MeSH terms included "pregnancy,prolonged" and "post$ pregnan$.tw".

Selection Criteria. Study designs considered included randomized controlled trials, cohort studies, and large ($n \geq 20$) case series with or without controls. Studies were included if the study population included women with prolonged pregnancy and data were provided that were relevant to one or more of the key research questions. Studies were excluded from formal abstraction if they did not report on original research, the patient population did not include women with prolonged pregnancy, the study design was a single case report or small case series, or a 2-by-2 table could not be constructed (for studies of test characteristics).

Data Collection and Analysis. Paired reviewers independently screened each abstract and article and performed the data abstraction. Included studies were graded for internal and external validity. Supplemental data were collected from the Nationwide Inpatient Sample.

Main Results. Although there is no direct evidence that antepartum testing reduces perinatal mortality in prolonged gestation, retrospective data suggest that morbidity may be reduced. Selection of appropriate outcomes for evaluating antepartum testing is difficult since mortality and morbidity are rare, and commonly used surrogate markers have substantial weaknesses. All currently used tests and combinations of tests have better specificity than sensitivity but good negative predictive values. There are no definitive data supporting the superiority of any particular testing method.

Most studies of interventions for the induction of labor do not report results specifically for women induced because of prolonged pregnancy or its complications. In general, agents that result in more efficient induction of labor also have higher rates of fetal heart rate pattern changes associated with frequent uterine contractions.

Pooled analysis of randomized trials of planned induction versus expectant management with antepartum testing suggests that planned induction reduces the risk of perinatal death with no increase in other perinatal or maternal morbidity, including cesarean section. At least 500 inductions are needed to prevent one perinatal death.

There are virtually no data on patient values and preferences for management options. There also are no published data on potential differences in epidemiology or outcomes of prolonged

pregnancy in racial, ethnic, or socioeconomic subgroups and no data allowing comparison of the cost-effectiveness of different strategies for managing prolonged pregnancy.

Conclusions. Induction of labor at 41 weeks or beyond results in fewer perinatal deaths compared with antepartum testing, but at least 500 inductions are necessary to prevent one death. There is insufficient evidence to recommend any specific induction agent in this setting. Additional high-quality research is needed.

Contents

Evidence Tables

Appendixes

Tables

Figure

Agency for Healthcare Research and Quality • www.ahrq.gov

Evidence Report/Technology Assessment

Number 53

Management of Prolonged Pregnancy

Summary

Overview

The estimated date of confinement, or due date, for normal pregnancies is calculated as 38 weeks after conception, or 40 weeks after the first day of the last normal menstrual period (assuming a "normal" 28-day menstrual cycle). Prolonged pregnancy has traditionally been defined as a pregnancy that extends 2 weeks or more beyond the estimated day of confinement, or 42 weeks. Approximately 18 percent of pregnancies in the United States extend beyond 41 weeks, and 7 percent extend beyond 42 weeks.

It has long been known that pregnancies extending many weeks beyond the average length are at increased risk for adverse outcomes, both because certain fetal anomalies, such as anencephaly, are associated with prolonged pregnancy, and also because of an increased incidence of stillbirth among otherwise normal infants. The increasing availability of ultrasound has significantly improved the accuracy of pregnancy dating and detection of fetal anomalies, so that extremely long gestations are rare. However, adverse outcomes continue to be associated with prolonged gestation.

In some cases, these risks appear to be due to uteroplacental insufficiency, resulting in eventual fetal hypoxia. Data from large registries show that the risk of perinatal death, especially of antepartum stillbirth, increases with advancing gestational age. If risk is calculated based on the number of ongoing pregnancies, gestational-age-specific stillbirth risk reaches a nadir at 37-38 weeks and then begins to increase slowly. Risks increase substantially after 41 weeks; however, the absolute risk is still low (between 1 and 2 per 1,000 ongoing pregnancies between 41 and 43 weeks).

Other adverse outcomes associated with uteroplacental insufficiency include meconium aspiration, growth restriction, and intrapartum asphyxia. In other cases, continued growth of the fetus leads to macrosomia, increasing the risk of labor abnormalities, shoulder dystocia, and brachial plexus injuries. Potential maternal risks associated with prolonged gestation, besides the obvious emotional trauma accompanying an unexpected fetal death or serious complication, include potential increased risk of injury to the pelvic floor associated with difficult deliveries of macrosomic infants. Interventions intended to prevent adverse perinatal outcomes, such as induction of labor and cesarean section, may themselves carry iatrogenic risks, such as increased rates of infection, hemorrhage, or other complications.

Several strategies currently are used in practice to prevent adverse outcomes associated with advancing gestation. Testing methods developed for reducing perinatal morbidity and mortality in women with high-risk pregnancies because of diabetes, hypertension, or other complications of pregnancy have been applied to women with pregnancies extending beyond 40 weeks. Another strategy, induction of labor at a predefined gestational age, has been proposed and evaluated as a method of reducing perinatal mortality and other adverse outcomes associated with prolonged gestation. However, because the point at which the risk of adverse outcomes outweighs the risks and costs of active interventions is uncertain, controversy remains about the optimal timing and

methods for managing increased risks to both fetus and mother associated with prolonged gestation.

Investigators at the Duke University Evidence-based Practice Center reviewed the evidence concerning the benefits, risks, and costs of commonly used tests, induction agents, and strategies for reducing the risks associated with prolonged gestation. Because of the inherent uncertainty in estimates of gestational age, variability in the length of otherwise uncomplicated pregnancies, and the lack of clear consensus on when risks of adverse outcomes outweigh risks of intervention, the researchers did not restrict the review to interventions performed only after a specified gestational age.

This summary and an evidence report were prepared based on the Duke EPC review. The primary target audiences for the summary and evidence report are groups involved in writing guidelines or educational documents on management of prolonged pregnancy for health care professionals. Secondary audiences include health care professionals providing care for pregnant women (obstetricians, family physicians, nurse-midwives, nurses, childbirth educators, etc.); policymakers involved in payment decisions; agencies involved in funding basic, clinical, and health services research; media involved in dissemination and education about health issues; and patients with an interest in reviewing the medical literature concerning management of prolonged pregnancy.

Reporting the Evidence

Key Research Questions

Four key research questions were addressed:

1. What are the test characteristics (reliability, sensitivity, specificity, predictive values) and costs of measures used in the management of prolonged pregnancy (a) to assess risks to the fetus and mother of prolonged pregnancy and (b) to assess the likelihood of a successful induction of labor?
2. What is the direct evidence comparing the benefits, risks, and costs of planned induction versus expectant management at various gestational ages?
3. What are the benefits, risks, and costs of currently available interventions for the induction of labor?
4. Are the epidemiology and outcomes of prolonged pregnancy different for women in different ethnic groups, socioeconomic groups, or age groups (i.e., adolescents)?

Interventions Assessed

The following interventions were considered:

Testing

1. Tests to determine risk of stillbirth or compromise related to prolonged gestation, including:
   - Maternal measurement of fetal movement.
   - Nonstress test (NST).
   - Contraction stress test (CST), using either nipple stimulation or oxytocin.
   - Amniotic fluid measurements: biophysical profile, using either five measures (reactive NST, breathing, tone, movement, amniotic fluid), or two measures (NST, amniotic fluid).
   - Doppler measurements of umbilical or fetal cerebral blood flow.
2. Tests to determine the risk of macrosomia, including estimation of fetal weight (maternal judgment, clinical examination, ultrasound).
3. Tests to estimate likely success of induction of labor, including:
   - Clinical estimation of cervical ripeness (Bishop score).
   - Fibronectin.

Management Options Other than Testing

1. No intervention (either induction or testing).
2. Interventions to prevent prolonged pregnancy (scheduled sweeping of membranes).
3. Planned induction (either 41 weeks, 42 weeks, or later).
4. Testing for fetal well-being (using tests described above):
   - Varied time of initiation (40, 41, 42 weeks).
   - Varied frequency.

Specific Agents/Interventions Used to Induce Labor

- Amniotomy
- Castor oil
- Extra-amniotic saline instillation
- Relaxin
- Sweeping of the membranes
- Foley catheter
- Nipple stimulation
- Oxytocin
- Prostaglandins (prostaglandin E2 gel, tablets, and inserts; misoprostol)
- Mifepristone

The researchers did not attempt to systematically review the basic and clinical research on the physiology of normal parturition, the role of routine ultrasound in early pregnancy, or interventions performed during labor and delivery to reduce the risks of adverse outcomes of conditions associated with, but not unique to, prolonged pregnancy (such as oligohydramnios or meconium-stained amniotic fluid).

Patient Population and Settings

The primary patient population considered in the review was pregnant women with a single fetus in the vertex position, approaching or past the estimated date of confinement, without any other medical or obstetrical complications (including prior cesarean section), where the only potential factor increasing the risk of an adverse perinatal or maternal outcome was advancing gestational age. The researchers also examined the potential interaction of this risk with age and race/ethnicity. The principal practice settings considered were hospitals, freestanding birthing centers, patients' homes, and prenatal clinics or other facilities where ambulatory prenatal care is delivered.

Outcomes Considered

Outcomes considered varied depending on the study and the question being addressed, but the researchers focused primarily on clinically relevant outcomes. Data recorded included anatomic outcomes (changes in cervical dilation or Bishop score); perinatal and maternal mortality; surrogate markers of fetal compromise (nonreassuring changes in fetal heart rate patterns, meconium); mode of delivery (cesarean, vaginal, operative vaginal); other interventions (need for labor augmentation, need for labor induction); adverse outcomes (complications of vaginal and cesarean delivery, complications of interventions); and use of resources (time to delivery, length of stay, medication, and labor costs).

Methodology

Literature Sources Used

The primary sources of literature were the following databases (with search years shown in parentheses) MEDLINE (1980-December 2000), HealthSTAR (1980-December 2000), CINAHL (1983-December 2000), Cochrane Database of Systematic Reviews (CDSR) (Issue 4, 2000; Issue 1, 2001; and Issue 2, 2001), Database of Abstracts of Reviews of Effectiveness (DARE), and EMBASE (1980-Jan 2000). Searches of these databases were supplemented by secondary searches of reference lists in all included articles, especially Cochrane review articles, scanning of current issues of journals not yet indexed in the computerized bibliographic databases, and suggestions from an advisory panel.

The initial searches were performed in MEDLINE and then duplicated in other databases. All searches were limited to English-language articles published since 1980 involving human subjects. The cut-off threshold of 1980 was based on the lack of general availability of ultrasound prior to that date. It was judged that trials conducted and published prior to 1980 would be problematic both in terms of the accuracy of diagnosis and comparability with current testing and management strategies. Primary MeSH terms used in all searches included "pregnancy,prolonged/" and "post$ pregnan$.tw."

Screening of Articles

The searches yielded 701 English-language articles. Abstracts from these articles were reviewed against the inclusion/exclusion criteria by six physician investigators, with assistance from one senior medical student. A team of two investigators reviewed each abstract; when no abstract was available, the title, source, and MeSH words were reviewed. At this stage, articles were included if requested by one member of the team. At the full-text screening stage, two investigators independently reviewed each article, and disagreements were resolved through discussion.

Each screened article was coded according to three topic areas: (a) testing: two or more tests were compared in terms of accuracy or agreement of test results, or the test result was correlated with some health outcome; (b) management: the article addressed the relative effectiveness of planned induction versus expectant management or the relative effectiveness of an induction agent; and (c) testing and management: some combination of the above.

Included study designs were determined by the article's topic area. Study designs for articles on testing or testing and management included randomized controlled trials, cohort studies, and large case series (at least 20 subjects). The only study design included for management articles was the randomized controlled trial.

Studies of these types were included if they met the following criteria:

- Study population included women with prolonged pregnancy.
- Study provided data relevant to at least one of the four key questions described above.

- Study reported health outcomes, use of health services, or economic outcomes related to the management of prolonged pregnancy.

Exclusion criteria included:

- Article was not original research.
- Article did not address prolonged pregnancy.
- Study design was a single case report.
- Study design was a small case series with fewer than 20 subjects.
- Article evaluated testing, but data provided were insufficient to construct 2-by-2 tables of test sensitivity and specificity.

Data Abstraction Process

Teams of two investigators performed the data abstraction for eligible articles identified at the full-text screening stage. For each included article, one physician completed the data abstraction form, and the other served as an "over-reader." The information from the data abstraction form—including details on study characteristics, patient population, outcomes, and quality measures—was then summarized into evidence tables. Data abstraction assignments were made based on clinical and research interests and expertise.

Criteria for Evaluating the Quality of Articles

Using criteria developed for prior evidence reports, the researchers evaluated each article for the presence or absence of factors influencing internal and external validity. These criteria were:

- For management articles: Randomized allocation to treatment and appropriate methods of randomization; adequate description of the patient population to allow comparison with the intended patient population, including descriptions in terms of gestational age, criteria used to assign gestational age, and measurement of baseline cervical ripeness; description of criteria used to make management decisions associated with primary outcomes such as cesarean delivery; and recognition and discussion of important statistical issues such as sample size and use of appropriate tests.
- For testing articles: The above criteria, plus description of an implicit or explicit reference standard, discussion of issues of verification bias, measurement of test reliability, and adequate description of the testing protocol.

Additional Data Sources

The researchers also examined discharge data from the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample maintained by the Agency for Healthcare Research and Quality. This database contains administrative discharge data from over 1,000 hospitals in 22 States (at the time of the review), representing a stratified sample of 20 percent of U.S. hospitals. The researchers used these data to provide supplemental information on differences in the epidemiology and outcomes of prolonged pregnancy between ethnic and socioeconomic groups. Using ICD-9 codes, they divided all deliveries into "preterm" (644.2x), prolonged (645.x), and "term" (all other delivery codes). The researchers examined differences in outcomes between coded ethnic groups (white, black, Hispanic, Asian/Pacific Islander, American Indian, and other) and by insurance status (Medicare, Medicaid, private/health maintenance organization, self-pay/no insurance, "no charge," and "other") within these categories.

Findings

The principal findings of the report are summarized here.

- The risk of antepartum stillbirth increases with increasing gestational age. Data from several large studies in the United Kingdom show that, when calculated as deaths per 1,000 ongoing pregnancies, antepartum stillbirth rates begin increasing after 40 weeks, with estimates of 0.86-1.08/1,000 between 40 and 41 weeks, 1.2-1.27/1,000 between 41 and 42 weeks, 1.3-1.9/1,000 between 42 and 43 weeks, and 1.58-6.3/1,000 after 43 weeks. Gestational-age-specific morbidity risks using the same methodology were not available.
- There is no direct, unbiased evidence that antepartum testing reduces perinatal morbidity and mortality in prolonged gestation. Retrospective data suggest higher risks of morbidity in women who did not receive testing, but it is unclear whether other factors contributed to these excess risks.
- As the sensitivity of antepartum testing for predicting surrogate markers of fetal compromise increases, specificity decreases. Testing strategies involving a combination of fetal heart rate monitoring and ultrasonographic measurement of amniotic fluid volume appear to have the highest levels of sensitivity. However, methodological issues and variability in specific tests and testing strategies prohibit definitive conclusions about which test or combination of tests has the best performance.

- Qualitatively, there is a consistent trend seen in studies of antepartum testing: test sensitivity is worse than test specificity, yet test-negative predictive values are greater than test-positive predictive values. This suggests that the high negative predictive values observed are because of an overall low risk of adverse outcomes. Unless test sensitivity increases with increasing gestational age (for which the researchers found no evidence), the negative predictive value will decline as gestational age advances, since the risk of adverse outcomes increases with advancing gestational age. Declining negative predictive values mean higher rates of false-negative antepartum tests and potentially higher rates of perinatal complications.
- Although the risk of antepartum stillbirth increases with increasing gestational age, there is no evidence that allows determination of the optimal time to initiate antepartum testing. Specifically, there is no evidence that testing prior to 41 weeks in otherwise uncomplicated pregnancies improves outcomes for either mother or infant.
- Both ultrasound and clinical assessment are reasonably sensitive in predicting birthweights greater than 4,000 grams in prolonged pregnancy, but they perform less well at predicting the more clinically relevant weight of greater than 4,500 grams. Evidence from one randomized trial shows that induction of labor based on estimated fetal weight does not improve outcomes for either infant or mother. There also is no evidence that an antepartum diagnosis of birthweight greater than 4,000 grams improves outcomes.
- Clinical examination of the cervix may help predict successful induction. However, individual components of the examination exhibit substantial inter- and intraobserver variability.
- Published data do not allow estimation of the cost-effectiveness of tests of fetal well-being.
- Although not statistically significant in most individual trials, there is a consistent finding that perinatal mortality rates are lower with planned induction at 41 weeks or later compared with expectant management, a finding confirmed by formal meta-analysis. Based on the observed absolute risk difference in the meta-analysis, at least 500 inductions are necessary to prevent one perinatal death. Whether this is an acceptable trade-off at either the policy or individual level is unclear.
- Other perinatal outcomes did not appear to differ significantly between induction and expectant management groups.
- Maternal outcomes did not differ between women managed with antepartum monitoring or with planned induction in the included studies. Specifically, overall rates of cesarean section did not differ, either globally or in subgroup analysis. Subgroup analysis of one large trial suggested this was due to very high rates of cesarean section in women managed with antepartum testing who were induced because of abnormal antepartum testing, reaching a predefined induction date, or other indications.
- Only one large trial reported costs. Based on 1992 costs and care provided, the study found that planned induction at 41 weeks was less expensive than expectant management with antepartum testing. However, because of significant changes in the technologies used and the economics of medicine in the interim, additional research is needed to better understand the cost implications of these two strategies.
- There is a remarkable lack of data on patient-oriented outcomes, such as quality of life or measures of patient preferences for different outcomes or for different processes to achieve those outcomes.
- Castor oil given at term appears to be effective in promoting labor, with a consistent side effect of maternal nausea; whether other outcomes of interest are affected is unclear. Conclusions about safety cannot be drawn.
- Manual nipple stimulation at term may promote labor, but effectiveness may depend on the protocol used and patient adherence to the protocol. Currently available data are insufficient to draw conclusions about either effectiveness or safety.
- Data on the safety and effectiveness of electrical breast stimulation as a method for inducing labor in prolonged gestation are inconclusive because of small sample size and a low proportion of subjects induced for an indication of prolonged pregnancy.
- Data on the safety and effectiveness of relaxin are limited, and no conclusions can be drawn.
- Sweeping of the membranes at or near term is effective in promoting labor and reducing the incidence of induction for prolonged gestation. There is no increase in adverse maternal outcomes.
- In general, there is a tradeoff between the effectiveness of induction agents in terms of achieving delivery and shortening the time to delivery, on the one hand, and risks of uterine tachysystole, hyperstimulation, and potential fetal compromise on the other. In increasing order of effectiveness, slow-dose oxytocin is followed by fast-dose oxytocin; PGE2 appears more effective than oxytocin; and misoprostol is more effective than PGE2. The heterogeneity of the patient populations in the published literature prohibits conclusions about the benefits and risks of these agents when used in the induction of labor in prolonged pregnancy, either for women induced electively or for

women with abnormal fetal surveillance. All studies were underpowered to detect differences in many important outcomes related to safety of induction agents.

- Mifepristone (RU-486) is consistently effective in reducing the time to labor and the time to delivery in women after 41 weeks. However, all three published trials reported nonsignificant trends toward higher rates of intermediate markers of fetal compromise, including abnormal fetal heart rate tracings and low Apgar scores.
- Data on costs associated with the use of different methods for induction are insufficient to allow conclusions about cost-effectiveness.
- The current published literature on the epidemiology and management of prolonged pregnancy does not provide information on the potential effects of race and ethnicity, socioeconomic status, or age on the incidence and outcomes of prolonged pregnancy.
- Based on administrative data, the proportion of deliveries occurring after 42 weeks does not appear to differ between ethnic groups, despite clear differences in the proportions delivering at earlier gestations.
- Based on administrative data, black women with prolonged pregnancy are more likely to have low birthweight infants than white or Hispanic women. Black women also are more likely to have diagnoses of intrauterine growth restriction and oligohydramnios during prolonged pregnancies.
- Based on administrative data, women with prolonged pregnancies who are on Medicaid or have no insurance are more likely to have growth restriction and oligohydramnios compared with women who have private insurance.

Future Research

Future research on the management of prolonged pregnancy should include the following:

- Biomedical research into the mechanisms controlling the initiation of normal labor, the interaction of uterine contractile forces and the pelvic floor, and other factors involved in the process of labor and vaginal delivery is needed.
- Estimates of the risk of perinatal morbidity and mortality in the United States need to be generated from a variety of complementary data sources. Ideally, an estimate of these risks by gestational age and in women without intervention can be generated and will inform future individual and policy decisionmaking.
- Research is needed into the most effective and efficient ways of determining gestational age during prenatal care.
- Surrogate markers for fetal compromise need to be identified that are less susceptible to bias and observer variability and more clinically relevant than current markers.
- Study designs for evaluating fetal testing need to minimize the effects of verification bias and avoid outcomes that may be influenced by the test results.
- Sample size estimates for studies of interventions to induce labor should be based on the power to detect clinically relevant outcomes. In particular, adequate power to determine safety is needed.
- Studies of interventions designed to induce labor should provide data on the benefits and risks of these interventions in women induced solely because of advancing gestational age and in women followed with antepartum testing because of prolonged gestation who are induced because of abnormal test results.
- Research is needed to identify markers that reliably and reproducibly predict the probability of successful induction.
- Appropriate statistical measures of central tendency and of significance testing should be used in studies of both testing strategies and induction interventions.
- Data on the medical and nonmedical costs associated with prolonged gestation and its management are needed. Research into economic outcomes should consider the effects of policy changes on issues such as staffing.
- Data on patient preferences for management strategies and outcomes are needed.

Availability of the Full Report

The full evidence report from which this summary was taken was prepared for the Agency for Healthcare Research and Quality (AHRQ) by the Duke Evidence-based Practice Center, Durham, NC, under contract number 290-97-0014. It is expected to be available in late spring 2002. At that time, printed copies may be obtained free of charge from the AHRQ Publications Clearinghouse by calling 800-358-9295. Requesters should ask for Evidence Report/Technology Assessment No. 53, *Management of Prolonged Pregnancy*. In addition, Internet users will be able to access the report and this summary online through AHRQ's Web site at www.ahrq.gov.

www.ahrq.gov

AHRQ Pub. No. 02-E012
March 2002

ISSN 1530-440X

Evidence Report

Chapter 1. Introduction

This report presents the results of a systematic review of the available evidence on the benefits, risks, and costs of different strategies for managing prolonged pregnancy to avoid adverse perinatal and maternal outcomes. It was prepared for the Agency for Healthcare Research and Quality by investigators at the Duke Evidence-based Practice Center, Durham, NC.

Background

The "normal" length of gestation has traditionally been defined as 40 weeks, or 280 days, after the first day of the last menstrual period. This figure is used to calculate the "estimated date of confinement" or "due date." Postterm pregnancy is defined by the American College of Obstetricians and Gynecologists (ACOG) as a gestation longer than 42 weeks, or 294 days, from the onset of the last menstrual period (Anonymous, 1997). It has long been recognized that the risk of adverse fetal outcomes, such as stillbirth, meconium aspiration, asphyxia, and the dysmaturity syndrome, is increased as gestational age progresses beyond 42 to 43 weeks (Mannino, 1988). However, the appropriate gestational age at which a pregnancy should be considered "high risk" for reasons of advancing gestation alone is unclear for several reasons. We discuss issues surrounding the concept of "normal" gestational age in this section, then review the data on risks associated with advancing gestational age.

Normal Variation versus Pathology

The mechanisms involved in the onset of normal labor in humans are a complex interaction between the fetus, placenta, uterus, and cervix. The fetal central nervous system may play a key role. Changes in circulating hormones produced by the placenta, such as progesterone, and in local production of prostaglandin and other cytokines, intercellular communication between uterine smooth muscle cells, and changes in extracellular matrix in both the uterus and the cervix are all important, but the exact cascade of events involved remains to be elucidated. Given this complexity, normal variability in the length of otherwise uncomplicated pregnancies should be expected. Most women who have prolonged gestation likely represent one extreme of normal variability in gestational age; in other women, or in specific pregnancies in an individual woman, the mechanisms involved in preparing for labor or signaling the onset of labor may differ.

The most recent ACOG review of the subject of "postterm" pregnancy cites estimates of 3-14 percent of all pregnancies (Anonymous, 1997). Estimates of the proportion of pregnancies delivering after 41 or 42 weeks are subject to variability because of variable accuracy in dating. Randomized trials of routine screening with ultrasound in the second trimester have consistently shown that routine screening reduces the proportion of women induced for prolonged pregnancy when compared with selective screening (Crowley, 2000). Since routine ultrasound screening is not the standard of care in the United States, population-based estimates will necessarily be subject to error. The most recent available data from birth certificates (1999) suggest that 39.6 percent of all deliveries in the United States occur at 40 weeks or beyond, 18.7 percent at 41 weeks or beyond, and 7.4 percent at 42 weeks and beyond (Ventura, Martin, Curtin, et al., 2000). Because these data include women who delivered prematurely, either through spontaneous preterm labor or because of other pregnancy complications, and women who were induced for

other reasons, the data cannot be used to estimate mean or median gestational age. Interestingly, the proportion of all births between 40 and 42 weeks is somewhat lower for black women compared with white or Hispanic women, reflecting the higher risk of preterm delivery in black women. However, the proportion of women delivering after 42 weeks is similar among all three ethnic groups. If errors in gestational dating are randomly distributed among the three groups, then this suggests that true "postterm" pregnancies may be due to true differences in the biological process initiating labor in these pregnancies, rather than representing the extremes of the distribution of normal gestational length.

Even the concept of "normal" pregnancy length is more complex than it first appears. One possibility is to define it as the mean, median, or mode for all pregnancies, perhaps stratified by parity and race, with some predefined range that captures the majority of the population. This value would inevitably be skewed by preterm deliveries, both spontaneous and induced for other complications; however, this length would still be "normal" in the sense that it conveys the expected length of the gestation for any woman at the beginning of the pregnancy. Since every woman has some nonzero risk of preterm delivery at the start of the pregnancy, "normal" length defined in this manner has some meaning.

Alternatively, "normal" length can be defined as the length of gestation in women who have uncomplicated pregnancies, labors, deliveries, and perinatal outcomes in the absence of any obstetric intervention. One could then divide pregnant women into three separate populations: (1) those with normal outcomes in the absence of intervention; (2) those requiring intervention and/or experiencing adverse outcomes associated with preterm delivery; and (3) those requiring intervention and/or experiencing adverse outcomes associated with late delivery. We did not identify any reports that characterized gestational length in this manner. Such an exercise might prove useful as an alternative method for discussing risks associated with prolonged gestation. In other words, most of the literature addresses the question: "Given gestational age, what is the likelihood of adverse outcomes?" Clinically, this is very reasonable. An alternative way to think about the problem when defining "normal" length of gestation is to ask the following two questions: "Given a good outcome without any intervention, what is the average gestational age?" And (for the two populations of preterm and term or later pregnancies): "Given an adverse outcome, what is the average gestational age?"

Errors in Dating

Menstrual Dates

Prior to the ready availability of ultrasound in the 1980s, estimation of gestational age based on menstrual dates alone was often inaccurate. For example, women who conceived soon after stopping oral contraceptives were more likely to have prolonged gestations in one series (Keng and Eng, 1982). Even with accurate recall of dates, there will be some variability in gestational age estimation because the 40-week estimate is based on an assumption of an "ideal" 28-day menstrual cycle, with ovulation on day 14. Because the follicular phase is often quite variable (ranging from 7 to 21 days), this assumption (upon which most gestational age calculators are based) will inevitably lead to some over- or underestimation of gestational age and can lead to errors in understanding the relationship between gestational age, birthweight, and pregnancy outcome (Gjessing, Skjaerven, and Wilcox, 1999).

Ultrasound

The availability of ultrasound in most sites in the United States has substantially improved the ability to estimate gestational age more precisely. Randomized trials of routine versus selective screening with ultrasound in the second trimester have consistently found a reduced incidence of induction of labor for prolonged pregnancy in the routine screening groups, presumably because of more accurate dating (Crowley, 2000). However, ultrasound itself has a nonnegligible degree of error. The error is approximately ± 1 week for scans done in the first trimester, ± 2 weeks for scans done in the second trimester, and ± 3 weeks for scans done in the third trimester (ACOG, 1997). Thus, even for women with early ultrasound dating, the "true" gestational age falls within a 14-day window of time; that is, some women with a recorded gestational age of 41 weeks will actually be 42 weeks, and some will actually be 40 weeks. In addition, because ultrasound dating is based on embryonic or fetal size, an association between size at the time of the ultrasound and later outcomes can create systematic bias in assessing gestational age-associated risk (Henriksen, Wilcox, Hedegaard, et al., 1995). For example, ultrasound dating will consistently overestimate the gestational age of larger than average fetuses. This early overestimation of gestational age could create a bias that would lead to an overestimation of the association of advanced gestational age and macrosomia. On the other hand, gestational age will be consistently underestimated for smaller than average fetuses. If some conditions that lead to low birthweight manifest themselves very early in pregnancy, then this will lead to an underestimation of the association of conditions associated with low birthweight and advancing gestational age.

The effects of uncertainty in dating pregnancy are not insignificant. Population-based estimates of the outcomes of pregnancy by gestational age, clinical trial data, and policy and clinical decisions based on these data are all dependent on the accuracy of the determination of gestational age.

The population of pregnant women with "prolonged" pregnancy thus likely represents at least two distinct groups:

1. Women in whom gestational age is overestimated because of the inherent error of all methods of dating.

2. Women whose pregnancies are correctly dated. Some of these women may represent the outer limits of normal variability. Others may have underlying defects in the mechanisms signaling the onset of labor.

It is likely that the risk of adverse outcomes varies among these groups. Many of the monitoring strategies discussed throughout this report are designed to identify fetuses at higher risk of adverse outcomes. The following section discusses the adverse outcomes associated with prolonged gestation, as well as the degree to which the risk of these outcomes is related to gestational age.

Burden of Illness: Risks Associated with Prolonged Pregnancy

Adverse fetal outcomes associated with advancing gestation can be divided into two categories:

1. Those associated with decreased uteroplacental function, resulting in oligohydramnios, reduced fetal growth, passage of meconium, asphyxia, and, potentially, stillbirth.

2. Those associated with continued normal placental function, resulting in continued fetal growth, with a subsequent increased risk of trauma during birth, including shoulder dystocia with possible permanent neurologic injury.

Adverse physical consequences to the mother resulting from prolonged gestation include those associated with increased fetal size, including an increased risk of short-term trauma to the pelvic floor, vagina, and perineum (as well as a possible longer-term risk of pelvic floor dysfunction), and postpartum hemorrhage. Interventions performed to reduce the risk of perinatal morbidity and mortality, such as induction of labor or cesarean section, have iatrogenic risks, such as infection, hemorrhage, and surgical injury. In addition, any adverse outcome for an infant will obviously have significant emotional impact on the mother.

Risk of Perinatal Mortality

The risk of perinatal death decreases with advancing gestational age until some point between 38 and 41 weeks, when it begins to increase again. The gestational age at which the risk begins to increase and the degree of risk involved have been subject to a reconsideration in several recent publications (Table 1). Yudkin, Wood, and Redman (1987) examined data from 40,888 deliveries in the Oxford Health District in England between 1978 and 1985. When unexplained stillbirth rates were calculated using the number of total deliveries within a given gestational age period, the rate per 1,000 births was 2.14 from 37 through 38 weeks, 0.43 from 39 through 40 weeks, and 1.24 from 41 weeks on. When estimated using a different denominator, the number of continuing pregnancies (i.e., the number of pregnancies still at risk of having a stillbirth), rates were different: 0.42/1,000 for 37 and 38 weeks, 0.29/1,000 for 39 and 40 weeks, and 1.24/1,000 for 41 weeks and later.

Hilder, et al., examined data from 171,527 births from the North East Thames Region in London (Hilder, Costeloe, and Thilaganathan, 1998). Stillbirth rates calculated as a percentage of all deliveries declined from 6.2/1,000 at 37 weeks to 1.5/1,000 at 40 weeks, then began to increase again with advancing gestational age (1.7 at 41 weeks, 1.9 at 42 weeks, and 2.1 at 43 weeks or more). The pattern was slightly different when risk was estimated as stillbirths per 1,000 ongoing pregnancies: 0.34 at 37 weeks, 0.70 at 38 weeks, 0.83 at 39 weeks, 1.57 at 40 weeks, 1.48 at 41 weeks, 3.29 at 42 weeks, and 3.71 at 43 weeks and beyond.

Cotzias, Paterson-Brown, and Fisk (1999) performed a reanalysis of the data set used by Hilder's group. In addition to estimating the number of stillbirths in a given gestational age divided by the number of ongoing pregnancies, the authors also estimated the "prospective

stillbirth risk," the total number of stillbirths at or beyond a given gestational age divided by the total number of pregnancies at or beyond that age, multiplied by 1,000. Other data sets were used to estimate the proportion of singleton births and the proportion of stillbirths occurring in singleton pregnancies, as well as the proportion of stillbirths that were unexplained by anomalies or other recognized fetal and maternal complications. Using this methodology, the risk for unexplained stillbirth in singleton pregnancies was highest at 37 weeks (1.55/1,000), declined to a low of 1.08/1,000 at 40 weeks, then increased again to 1.58/1,000 at 43 weeks. The high rates at lower gestational ages may reflect this methodology.

Most recently, Smith (2001) analyzed data from Scotland for the period 1985 through 1996. This analysis has several advantages over the previous ones. First, the number of deliveries is considerably larger, resulting in greater precision of risk estimates. Second, stillbirths are divided into antepartum and intrapartum stillbirths, a distinction that has clinical relevance, since clinical strategies for preventing each of these might be quite different. Third, congenital anomalies were explicitly excluded. Fourth, life table methods were used to account for censoring resulting from deliveries within a given observation period. Fifth, the time period is considerably later, making the results more likely to reflect current clinical management, at least in the United Kingdom. Finally, cumulative probabilities for stillbirth at each gestational age were estimated.

Estimates of antepartum stillbirth in this paper show the conditional probability increasing as gestational age increases (Table 1), while the probability of intrapartum stillbirth does not change significantly with increasing gestational age. Smith (2001) also found that cumulative probability increases, from 0.4/1,000 at 37 weeks to 2.2 /1,000 at 40 weeks to 11.5/1,000 at 43 weeks. The risk of any perinatal death, when calculated as a cumulative probability, begins to increase at 39 weeks; when calculated as a risk per total births in a given week, it does not begin to increase until after 42 weeks. Risks did not appear to differ when deliveries between 1985 and 1990 were compared with those between 1991 and 1996; however, risks for antepartum stillbirth were increased significantly for primigravidas compared with parous women.

The advantage of cumulative probability is that it captures the risk of death in preceding gestational ages. Smith (2001) uses the metaphor of Russian roulette to explain the difference between conditional probability and cumulative probability: the risk with each pull of the trigger is 1 in 6, but the risk of death for someone taking his fifth shot is greater than for someone taking his first shot. For example, Smith estimated the conditional probability of stillbirth at 43 weeks as 6.3/1,000 ongoing pregnancies, while the cumulative probability was 11.5/1,000 ongoing pregnancies. This difference represents the effects of stillbirths occurring before 43 weeks. The potential clinical significance of this is that achieving the absolute minimum cumulative stillbirth probability may require interventions at earlier gestational ages.

Consistently, the risk of stillbirth in the above-described studies rises with advancing gestational age, and this increase appears to begin at 39-40 weeks when estimated using the number of ongoing pregnancies as the denominator. One limitation of these studies is that they were all performed in the United Kingdom, and the degree to which the risks would differ in a different population with different clinical management is unclear. Another limitation is that other potential causes of perinatal mortality, such as maternal diabetes or hypertension, are not explicitly accounted for in these data sets. Also, autopsy verification that fetal anomalies or other anatomic causes of death did not occur was not performed. However, a recent Norwegian case-control study of unexplained stillbirth, in which autopsy verification was performed and logistic regression was used to control for documented maternal disease, found that increasing gestational age remained a significant risk factor for unexplained stillbirth, along with maternal

age, smoking, obesity, and low educational level. Interestingly, parity was not a risk factor in the multivariate analysis (Froen, Arnestad, Frey, et al., 2001).

It should be pointed out that the risk of stillbirth in these studies remains quite low at an absolute level. The point at which the risk becomes unacceptable and justifies intervention is unclear and is likely to be influenced by each couple's feelings about the tradeoffs between intervention and no intervention.

Two other studies provide additional indirect evidence of increased risk of death with prolonged gestation. Bastian, Keirse, and Lancaster (1998) compared outcomes of all planned home births in Australia from 1985 through 1990 with all Australian births in the same time period and home births in other countries. The planned home birth perinatal death rate was 6.4/1,000 (46/7,002 total home births). Of the 44 deaths with known gestational age, seven (15.9 percent) were greater than 42 weeks. On chart review, six of these deaths, or 28.6 percent of the total, were classified as due to intrapartum asphyxia; prolonged pregnancies represented 10.7 percent of all home births. Overall, the mortality rate for home births in infants over 42 weeks was twice that for other home births. The authors point out that other conditions associated with perinatal mortality are much less common in the home-birth population, so that the excess mortality observed is unlikely to be solely due to the confounding effects of other complications, such as preeclampsia or diabetes.

Mehl-Madrona and Madrona (1997) reviewed self-reported data from midwives in the western United States between 1970 and 1985. A total of 4,361 midwife-attended home births were compared with 4,107 family-practitioner-attended home births performed in California and Wisconsin during the same time period. Sampling frames and response rates were variable, as were the data collection instruments. Deliveries were matched by maternal age, insurance status, parity, and presence of risk factors. Midwives were significantly more likely to deliver postdate pregnancies, defined as gestational age greater than 42 weeks, than were family practitioners (midwives also were more likely to deliver breech and twin pregnancies). Mortality rates were significantly higher for midwives compared to family practitioners, a difference that was attributable entirely to more postdate, twin, and breech deliveries in the midwife group.

Both of these studies are limited by issues concerning accuracy of dating, completeness of reporting, confirmation of causes of death, and in the case of the Mehl-Madrona paper, a rather complicated sampling scheme and questions about the true comparability of groups. There also are concerns about generalizability in terms of current midwifery practice in the United States. However, patients who select home birth are, by definition, low-risk patients. They also are unlikely to have undergone antepartum testing. The excess mortality seen in women with prolonged pregnancy delivering at home in these two studies is consistent with an independent effect of increasing gestational age on perinatal mortality.

Causes of Perinatal Mortality in Prolonged Pregnancies

Analysis of data from the Medical Birth Registry of Norway from 1978 to 1987 found that the risk of perinatal death was over five times higher in infants below the 10^{th} percentile of birthweight for their gestational age (odds ratio [OR], 5.68; 95 percent confidence interval [CI], 4.37 to 7.38) than in infants from the 10^{th} to 90^{th} percentile (Campbell, Ostbye, and Irgens, 1997), after adjustment for a variety of potential confounding variables, such as maternal complications like diabetes. Maternal age ≥ 35 years was also a risk factor in multivariate analysis (OR, 1.88; 95 percent CI, 1.22 to 2.89). Infants above the 90^{th} percentile in weight had a

decreased mortality risk (OR, 0.51; 95 percent CI, 0.26 to 1.00). A similar relationship between perinatal mortality in prolonged pregnancy and low birthweight was found in a review of Swedish registry data from 1987 through 1992 (Divon, Haglund, Nisell, et al., 1998). These observations are consistent with a hypothesis that decreased uteroplacental function, leading to growth restriction, oligohydramnios, and eventually asphyxia, is one of the major risks of advancing gestational age, although changes in weight occurring after death and prior to delivery may explain some of this phenomenon. What is not clear is whether the decreasing uteroplacental function is an inevitable result of advancing gestational age, or whether failure to go into labor is somehow a marker for some forms of uteroplacental insufficiency.

The Norwegian data are limited by the population (results may not be generalizable to a more diverse U.S. population), accuracy of dating (gestational age in the registry is based on last menstrual period), and time (obstetric management has changed somewhat since 1987). However, the observed association between low birthweight and perinatal mortality in a genetically homogeneous population with a relatively high standard of living and level of access to prenatal care suggests that this is at least partly a reflection of changes in the biology of the uterus, placenta, and/or fetus associated with prolonged pregnancy.

Another issue that should be considered in reviewing recent population-based data on perinatal mortality is the degree to which observed perinatal deaths are preventable. It is unclear from population-based administrative data what proportion of unexplained stillbirths after 40 weeks gestation occurred in women undergoing some form of antenatal surveillance. This information is important for two reasons. First, in order to estimate the benefits of antenatal surveillance at different gestational ages quantitatively, the baseline gestational-age-specific risk, in the absence of surveillance, is needed. Second, if current mortality data reflect mostly women who are undergoing surveillance, then the limits of currently available technology may have been reached; in this case, the only strategy available for further reducing perinatal mortality would be elective induction of labor at a predefined gestational age. This is supported by the findings of a Cochrane meta-analysis (Crowley, 2000), which showed an excess of perinatal mortality in the testing arms. Conversely, if current mortality data reflect women who are not undergoing surveillance, then greater efforts are needed to ensure access to currently available technologies.

Perinatal Morbidity

In the Norwegian database, risks for fetal distress in labor (relative risk [RR], 1.68; 95 percent CI, 1.62 to 1.72) and shoulder dystocia (RR, 1.31; 95 percent CI, 1.21 to 1.42) were significantly increased in infants born after 42 weeks compared with infants born between 39 and 42 weeks (Campbell, Ostbye, and Irgens, 1997). Others also have noted an association between prolonged pregnancy and increased fetal weight and/or shoulder dystocia (Acker, Sachs, and Friedman, 1985; Eden, Seifert, Winegar, et al., 1987; Nocon, McKenzie, Thomas, et al., 1993; Sarno, Hinderstein, and Staiano, 1991).

Data on longer term outcomes of infants born after prolonged gestations are relatively sparse. One Irish case-control study reported an association between prolonged pregnancy and neonatal seizures (Curtis, Matthews, Clarke, et al., 1988). In a study of British children with cerebral palsy, there was a strong association between maternal gestational age greater than 41 weeks and the presence of neonatal encephalopathy (defined as having both signs of neonatal neurological abnormalities and depression at birth, defined as a 1-minute Apgar score less than 6) (OR, 3.5;

95 percent CI, 1.0 to 12.1). This risk was particularly marked in primigravid women (OR, 11.0; 95 percent CI, 1.5 to 102.5). The infants studied also were more likely to have had induction of labor (indications not specified), long second stage of labor, meconium-stained amniotic fluid, and emergent cesarean section or operative vaginal delivery.

On the other hand, prospective studies have not shown an association between prolonged pregnancy and adverse physical or mental development at 1 or 2 years, even when stratified by presence or absence of the dysmaturity syndrome (Shime, Librach, Gare, et al., 1986).

In summary, available data are insufficient to quantify the degree of excess risk, if any, of perinatal morbidity (including neurological morbidity) associated with prolonged pregnancy.

Maternal Outcomes

Maternal risks of obstetric trauma and hemorrhage are increased in prolonged pregnancy compared with term pregnancy (Campbell, Ostbye, and Irgens, 1997). Labor abnormalities also are increased. All three of these may be related to an increased risk of macrosomia. Another potential reason, as stated above, is that some women who do not go into labor within the "normal" length of gestation have differences in the physiology of labor and delivery compared with women who begin labor earlier in gestation.

Interventions performed to prevent adverse outcomes associated with prolonged gestation have the potential for complications, most notably hyperstimulation resulting from too frequent uterine contractions, infection, bleeding, or organ injury from cesarean section.

Summary: Risks of Prolonged Pregnancy

Prolonged gestation is associated with an increased risk of perinatal death, as well as perinatal morbidities related to either uteroplacental insufficiency or fetal macrosomia. Direct maternal risks are potentially related to fetal macrosomia or to interventions used in the management of prolonged pregnancy. The gestational age at which the risk of adverse direct perinatal or maternal outcomes justifies the costs and potential complications of active intervention is unclear.

Scope and Purpose

The purpose of this evidence report is to review the evidence regarding strategies to reduce the risks of adverse maternal and fetal outcomes associated with advancing gestational age. Because of the issues discussed above, we did not limit our review to interventions performed after a predefined gestational age cut-point. Although "postterm" pregnancy technically refers to gestations beyond 42 weeks, and "postdate" to pregnancies beyond 40 weeks, others have used the phrase "prolonged pregnancy." The appropriate gestational age range upon which this report should focus proved a lively topic for debate among the members of the project's advisory panel of technical experts. However, consensus was reached that the primary focus should be on managing those risks associated with advancing gestational age, with an attempt at quantifying the gestational-age-specific risk. Because of this scope, we use the term "prolonged pregnancy" throughout this report, to avoid confusion with terminology associated with specific gestational age definitions. We use "postterm" and "postdate" only when specifically referred to in articles under discussion.

There is an inherent uncertainty associated with any estimate of gestational age. However, risks of certain adverse outcomes for both mother and infant clearly increase as gestational age increases after 37-38 weeks. Strategies to minimize these risks may themselves carry certain risks. The ultimate goal of this report is to provide a framework for rationally comparing these competing risks, and to help patients, clinicians, and policymakers decide for themselves the best options for managing prolonged gestation in their particular situation.

Key Research Questions

The key research questions addressed in the report were developed by the Agency for Healthcare Research and Quality (AHRQ) and our report partner, ACOG, and refined in consultation with AHRQ, ACOG, and the project's advisory panel of technical experts. The questions were as follows:

1. What are the test characteristics (reliability, sensitivity, specificity, predictive values) and costs of measures used in the management of prolonged pregnancy to (a) assess risks to the fetus and mother of prolonged pregnancy, and (b) assess the likelihood of a successful induction of labor?

2. What is the direct evidence comparing the benefits, risks, and costs of planned induction versus expectant management at various gestational ages?

3. What are the benefits, risks, and costs of currently available interventions for the induction of labor?

4. Are the epidemiology and outcomes of prolonged pregnancy different for women in different ethnic groups, different socioeconomic groups, or in adolescent women? This question reflects AHRQ's programmatic interest in identifying health disparities attributable to age, race/ethnicity, and socioeconomic status.

Our approach to addressing each of these questions was to identify and evaluate the relevant literature and supplemental data (if any); report the results; and where evidence was lacking or methodological limitations in the available sources precluded drawing firm conclusions, identify the issues needing resolution in order to answer the question.

Because the primary focus of the report is on clinical issues surrounding advancing gestational age, we did not systematically review the basic science literature on the initiation of labor, the physiology of the gravid uterus and cervix, placental function, or any of the other topics critical to a comprehensive understanding of these issues. The Duke team, AHRQ, ACOG, and the advisory panel all agreed that the time, effort, and additional expertise required to systematically review this literature precluded their inclusion in this evidence report.

Interventions Assessed

Based on the key research questions, our preliminary review of the literature, and discussions with the advisory panel, we considered the following interventions to reduce risks to the fetus or mother associated with advancing gestational age.

1. Testing:
   a. Tests to determine risk of stillbirth or compromise related to prolonged gestation:
      - Maternal measurement of fetal movement.
      - Nonstress test (NST).
      - Contraction stress test (CST), using either nipple stimulation or oxytocin.
      - Amniotic fluid measurements.
      - Biophysical profile, using either five measures (reactive NST, breathing, tone, movement, amniotic fluid) or two measures (NST, amniotic fluid).
      - Doppler measurements of umbilical or fetal cerebral blood flow.

   b. Tests to determine the risk of macrosomia.
      - Estimation of fetal weight:
        - Maternal judgment.
        - Clinical examination.
        - Ultrasound.

   c. Tests to estimate likely success of induction of labor.
      - Clinical estimation of cervical ripeness (Bishop score).
      - Fibronectin.

After discussion with the advisory panel, we did not include tests of fetal well-being that are no longer in widespread clinical use, such as estriol.

2. Management options other than testing:
   - No intervention (neither induction nor testing).
   - Interventions to prevent prolonged pregnancy:
     - Scheduled sweeping of membranes.
   - Planned induction:
     - 41 weeks.
     - 42 weeks.
     - Later timing
   - Testing for fetal well-being (using tests described above):
     - Varied time of initiation (40, 41, 42 weeks).
     - Varied frequency.

3. Specific agents/interventions used for the induction of labor:
   - Amniotomy.
   - Castor oil.
   - Extra-amniotic saline instillation.
   - Relaxin.
   - Sweeping of the membranes.
   - Foley catheter.
   - Nipple stimulation.
   - Oxytocin.
   - Prostaglandins:

  - Prostaglandin E_2 (gel, tablets, and inserts).
  - Misoprostol.
- Mifepristone.

We did not systematically review certain other interventions that may play a role in managing prolonged pregnancy. Although we discuss the effect of ultrasound estimation of gestational age on the diagnosis of prolonged pregnancy above, we did not attempt to systematically review the literature on the other potential benefits, risks, and costs of routine ultrasonography in early pregnancy. Attempting to place the potential benefits of accurate gestational dating for managing advancing gestational age in the context of the other possible outcomes associated with routine ultrasound screening was well beyond the scope of the report and beyond the resources available. Similarly, we did not systematically review the literature on intrapartum interventions used in the management of common complications of prolonged pregnancy (such as oligohydramnios or meconium-stained amniotic fluid) unless identified articles clearly included data on prolonged pregnancy.

Patient Populations

The primary patient population considered in this report was pregnant women with a single fetus in the vertex position, approaching or past the estimated date of confinement, without any other medical or obstetrical complications, where the only potential factor increasing the risk of an adverse perinatal or maternal outcome was advancing gestational age. We also examined the potential interaction of this risk with age and race/ethnicity. Our findings are specifically not applicable to women with prior cesarean section, for several reasons:

- Prior cesarean section was an exclusion criteria in the vast majority of the randomized trials of management strategies and induction agents; thus, we are unable to generalize these results.

- Recent observational data (Blanchette, Nayak, and Erasmus, 1999; Lydon-Rochelle, Holt, Easterling, et al., 2001; Plaut, Schwartz, and Lubarsky, 1999) suggest that risk of uterine rupture is increased in women with prior cesarean section undergoing induction of labor, especially with prostaglandins. Incorporating an evaluation of this evidence into the report would have required an additional consideration of the general risks and benefits of vaginal birth after cesarean section, which is well beyond the scope of this report.

Practice Settings

Practice settings where the interventions discussed in this report may potentially be considered for use include:

- Hospitals.
- Free-standing birthing centers.
- Patients' homes.
- Prenatal clinics or other facilities where ambulatory prenatal care is delivered.

Target Audiences

The primary target audiences for the evidence report are groups involved in writing guidelines or educational documents on management of prolonged pregnancy for health care professionals. Secondary audiences include:

- Health care professionals providing care for pregnant women (obstetricians, family physicians, nurse-midwives, nurses, childbirth educators, etc.).
- Policymakers involved in coverage/payment decisions.
- Agencies, foundations, and other groups involved in funding research.
- Media involved in dissemination and education about health issues.
- Patients with an interest in reviewing the state of the art of the medical literature concerning management of prolonged pregnancy.

Chapter 2. Methodology

In this chapter, we describe the basic methodology used to develop the evidence report, from topic assessment and refinement through the literature search, screening, and data abstraction process. Included are descriptions of the literature search strategies and results, literature sources, screening and grading criteria, quality control procedures, and supplemental data sources.

Topic Assessment and Refinement

A national advisory panel of technical experts was convened to work with the Duke research team. The 11-member panel included representatives from obstetrics-gynecology, including maternal-fetal medicine; pediatrics; childbirth education; and midwifery. In addition to the American College of Obstetricians and Gynecologists (ACOG), other major interest organizations represented on the panel included the American College of Nurse Midwives and the Adolescent Pregnancy Prevention Coalition of North Carolina.

Prior to our first conference call, the advisory panel and the Task Order Officer at the Agency for Healthcare Research and Quality (AHRQ) received a document that summarized the incidence and prevalence of prolonged pregnancy, described the characteristics and size of the affected population, identified the most affected practice settings and providers, specified the interventions to be considered, and presented a diagram of the conceptual model/causal pathway. The panel also received the four key questions specified in the task order. Based on Duke's preliminary assessment of the literature and discussion with the advisory panel and AHRQ Task Order Officer, all parties agreed to refine the key questions as follows:

1. What are the test characteristics (reliability, sensitivity, specificity, predictive values) and costs of measures used in the management of prolonged pregnancy to assess: (a) risks to the mother and fetus of prolonged pregnancy and (b) the likelihood of a successful induction?

2. What is the direct evidence comparing the benefits, risks, and costs of planned induction versus expectant management at various gestational ages?

3. What are the benefits, risks, and costs of currently available interventions for induction of labor?

4. Are the epidemiology and outcomes of prolonged pregnancy different for women in different ethnic groups, different socioeconomic groups, or in adolescent women?

In addition to reaching consensus on the key questions, the advisory panel agreed on the patient population, practice settings, and target audiences of the report, as described in Chapter 1 of this report. The causal pathway is represented in Figure 1.

Literature Search and Selection

The comprehensive review of the literature, from identification of databases through abstraction of individual articles into evidence tables, was a multi-step, sequential process.

Literature Sources

The primary sources of literature were six of the most widely used computerized bibliographic databases: MEDLINE (1980-December 2000), HealthSTAR (1980-December 2000), CINAHL (1983-December 2000), the Cochrane Database of Systematic Reviews (CDSR) (Issue 4, 2000; Issue 1, 2001; and Issue 2, 2001), the Database of Abstracts of Reviews of Effectiveness (DARE), and EMBASE (1980-Jan 2000). Searches of these databases were supplemented by secondary searches of reference lists in all included articles, especially Cochrane review articles, and scanning of current issues of journals not yet indexed in the computerized bibliographic databases. Titles regularly scanned included the *American Journal of Obstetrics and Gynecology*, the *British Medical Journal*, the *British Journal of Obstetrics and Gynaecology*, the *European Journal of Obstetrics and Gynecology and Reproductive Medicine*, the *International Journal of Gynecology and Obstetrics*, the *Journal of the American Medical Association*, the *Journal of Maternal-Fetal Medicine*, the *Journal of Obstetrics and Gynaecology*, *Obstetrics and Gynecology*, the *Lancet*, and the *New England Journal of Medicine*. Suggestions regarding search terms and specific articles were solicited from the advisory panel during two conference calls in December 2000 and March 2001 and resulted in additions to the literature database.

Search Strategy

We developed the basic search strategies using the National Library of Medicine's MeSH key word nomenclature developed for MEDLINE. The same strategies were used to search HealthSTAR and CINAHL. A Duke University Medical Center librarian checked the strategies and assisted with their translation to the key word structure used by EMBASE. Dr. Evan Myers searched the CDSR and DARE using "postterm pregnancy," "prolonged pregnancy," and similar terms.

The initial searches were performed in MEDLINE and then duplicated in other databases. All searches were limited to articles published since 1980, in the English language, and with human subjects. The cut-off threshold of 1980 was based on the general unavailability of ultrasound prior to that date. It was judged that trials conducted and published prior to 1980 would be problematic both in terms of the accuracy of diagnosis and comparability with current testing and management strategies. The decision to restrict the literature search to articles published since 1980 was agreed to by the members of the advisory panel.

The search strategies are reproduced in Tables 2 and 3.

Screening Criteria

Inclusion and exclusion criteria were developed for the literature searches so that the yield of articles would be appropriately focused. Empirical studies or review articles were excluded after screening based on the following criteria:

- Article was not original research.
- Article did not address prolonged pregnancy.
- The study design was a single case report.

- The study design was a small case series with fewer than 20 subjects.

Each screened article was coded as addressing one of three topic areas:

1. Testing: Two or more tests were compared in terms of the accuracy or agreement of test results or the test result was correlated with some health outcome.

2. Management: The article addressed the relative effectiveness of planned induction versus expectant management or the relative effectiveness of an induction agent.

3. Testing and management: Some combination of the above.

The criteria used to include articles were:

- The study population must address prolonged pregnancy; ideally, results should be reported separately for patients with prolonged pregnancy. Because it is possible that the response of the cervix and uterus to induction agents would be quite different in different clinical scenarios (both in terms of labor patterns and potential maternal and fetal side effects), studies of induction agents that did not include any otherwise healthy women with prolonged pregnancy were excluded.

- All original research or relevant reviews must relate to at least one of the four key questions described above.

- Outcomes were included if they were health outcomes or health services use or economic outcomes related to the management of prolonged pregnancy.

- We included only randomized controlled trials (RCTs) which used active or nonactive (i.e., placebo) controls for studies involving management topics. For testing articles, we included RCTs and those cohort and large case series that allowed construction of 2-by-2 tables for estimation of sensitivity and specificity. Articles that did not meet these criteria were not necessarily excluded from the review and often provided valuable background material. However, only articles meeting the inclusion criteria were formally abstracted into evidence tables.

Included study designs were determined by the article's topic area. Study designs initially included for testing articles and testing and management articles were case reports; small case series (< 20 subjects); medium to large case series (≥ 20 subjects); nonrandomized comparison studies (cohort or case series that used historical or concomitant nonrandomized controls); and RCTs. The study design of each screened article was coded in our literature database.

For the testing articles and testing and management articles, an evidence table entry was developed for each RCT and for each cohort study or large case series for which a 2-by-2 table linking test results to important outcomes could be constructed (Evidence Table 1). The only study design considered for management articles was the RCT. Our experience in past evidence report projects in which lack of data from RCTs necessitated the evaluation of nonrandomized studies has been that drawing inferences about the effectiveness of therapeutic interventions based on

nonrandomized studies is difficult, if not impossible, because of numerous biases and lack of consistency in data provided about important confounding variables. An evidence table entry was developed for each included management trial (Evidence Tables 2 and 3).

Screening Results

The literature searches yielded 701 English-language articles. A summary of the number of articles retrieved from each data source is provided in Table 4. The titles and abstracts of these articles were reviewed against the inclusion/exclusion criteria by seven investigators, Drs. Richard Blumrick, Elizabeth Livingston, Andrea Lukes, David Matchar, Douglas McCrory, and Evan Myers and a third-year medical student, Ms. Andrea Christian. Two investigators reviewed each citation. Abstracts were available for more than three-fourths of the citations; when no abstract was available, the title and source were screened. At this stage, articles were included if requested by one member of the review team. The full text of each article passing the title-and-abstract screen was retrieved from the library for further review.

At the full-text screening stage, each article was independently reviewed by two investigators, who forwarded their decisions to Ms. Jane Kolimaga, the task order manager, for recording and comparison. If indicated, reviewers were asked to reconcile differences of opinion. Overall, the teams initially disagreed on about 25-35 percent of their decisions, and all disagreements were resolved by consensus. In the event that two investigators could not agree, Dr. Evan Myers, the principal investigator, was to be the arbiter, but this situation never arose.

The task order manager coded the records in the bibliographic database at each screening stage. A summary of the results of the title-and-abstract and full-text screenings is provided in Table 5.

Data Abstraction

Teams of two investigators performed the data abstraction for eligible articles identified at the full-text screening stage: one performed the primary data abstraction, and the second “over-read” the abstracted information. A data abstraction form was developed prior to initiation of the formal abstraction process. During the development of the form, draft forms were reviewed by the investigators and Dr. Rebecca Gray, a nonclinician abstractor/editor, for clarity and completeness; as the person who converted the abstraction forms into evidence tables, Dr. Gray helped to insure that all relevant information was captured. The two final iterations of the form were pretested by the investigators who used them to abstract relevant data from a sample article. The information from the data abstraction form was then summarized in evidence table format by Dr. Gray. The data abstraction assignments were made by Dr. Myers based on the investigators’ clinical interests (e.g., management vs. testing). Copies of the data abstraction form and the evidence table template are provided in Appendixes 1 and 2, respectively.

Outcomes recorded included:

- Direct health outcomes:
  - Maternal mortality.
  - Perinatal mortality.

- Maternal morbidity (specific measures varied between studies; included infection, hemorrhage, perineal trauma, etc.).

- Perinatal morbidity (meconium aspiration, postmaturity syndrome, shoulder dystocia, brachial plexus injury, admission to neonatal intensive care unit).

♦ Surrogate measures:

- Neonatal umbilical artery pH, Apgar scores, meconium-stained amniotic fluid, nonreassuring fetal heart rate tracing.

- Cesarean section rates, overall and by specific indication.

♦ Resource use:

- Costs.

- Time to delivery, proportion of vaginal deliveries within a prespecified time.

♦ Test operating characteristics:

- Sensitivity, specificity, positive and negative predictive values for outcomes listed above.

Quality Scoring

We evaluated each study included in the evidence tables for factors affecting internal and external validity. For management articles, the elements of the quality scale were as follows:

♦ Were patients randomly assigned to the intervention?

♦ Was the method for randomization described, and if so, was it one shown to be associated with less bias (sealed envelopes) than others (alternating date or medical record number)?

♦ Was the patient population similar to the likely patient population?

♦ Were the intervention protocols clearly described or referenced?

♦ Were the criteria used to make management decisions associated with primary outcomes (such as cesarean section) described?

♦ Statistical issues: Were sample size and power issues discussed? Were the statistical tests used appropriate for the types of data analyzed?

♦ Was the study population described in terms of:
- Gestational age?

– Criteria used to assign gestational age?
– Bishop score or other measure of cervical ripeness?

For testing articles, we used the above criteria plus:

- Was an implicit or explicit reference standard defined?
- Was the issue of possible verification bias (patients with positive test results more likely to receive the reference standard test or treatment) addressed?
- Test reliability/variability: Was inter- or intrarater reliability of the test addressed?
- Was the study population well characterized in terms of the absence of risk factors such as diabetes, hypertension, etc.?
- Was the testing protocol described in sufficient detail to allow others to replicate it?

Scores on individual quality criteria were not aggregated into an overall score but were considered and reported individually. We preferred this approach for several reasons:

1. Previous work has shown that aggregated numeric scoring systems may not discriminate well between "high" and "low" quality studies, even for randomized trials (Jüni, Witschi, Bloch, et al., 1999; Moher, Jadad, and Tugwell, 1996).

2. Development and use of a new quality score would have required additional work for validation.

3. Identification of specific weaknesses in each study will be helpful in identifying trends, which in turn will assist with our recommendations for future research.

Our approach of describing key design components, rather than assigning a single aggregate score, is also consistent with recent recommendations from an expert panel on meta-analysis of observational studies (Stroup, Berlin, Morton, et al., 2000) and a recent review of the methodology of systematic reviews (Jüni, Altman, and Egger, 2001).

Summaries of the quality evaluation are provided in the evidence table entry for each abstracted article. A "+" indicates that a given criterion was met, a "-" signifies that the criterion was not met. The "+" and "-" notations were assigned by the primary abstractor and confirmed by the over-reader.

Quality Control Procedures

We employed quality-monitoring checks at every phase of the literature search, review, and data abstraction process to reduce bias, enhance consistency, and check the accuracy of screening:

- Medical librarian review of the literature search strategy.

- Review of literature search strategies by the advisory panel of technical experts.
- Check on completeness of the literature search results through reference list checks by the screener of each article.
- Reconciliation of all differences of opinion by reviewers on all full-text articles.
- Agreement of two reviewers for all eligible studies.
- Data abstractions completed by one investigator and reviewed (over-read) by another.
- Additional checks of evidence table entries for completeness and accuracy by a nonphysician abstractor.
- Solicitation of advice at key decision points from the advisory panel of technical experts.

Supplemental Data Sources

In order to get additional information about possible racial and socioeconomic differences in the incidence and outcomes of prolonged pregnancy, we analyzed data from the 1997 Nationwide Inpatient Sample (NIS) (Nationwide Inpatient Sample [NIS], 1997). The NIS is part of AHRQ's Healthcare Cost and Utilization Project (HCUP) and collects discharge data from a stratified sample of approximately 20 percent of U.S. hospitals. Using ICD-9 codes, we divided all deliveries into "preterm" (644.2x), prolonged (645.x), and term (all other delivery codes). We examined differences in outcomes between coded ethnic groups (white, black, Hispanic, Asian/Pacific Islander, Native American, and "other") and by insurance status (Medicare, Medicaid, private/health maintenance organization, self-pay/no insurance, "no charge," and "other") within these categories.

Supplemental Analyses

At the start of every evidence report project, we evaluate the feasibility of and need for meta-analyses, decision analyses, cost-effectiveness analyses, or a combination of all three. A decision about whether to proceed with such analyses is made based on the key questions and the state of the literature, after discussion with AHRQ and the advisory panel. We decided not to perform any supplemental analyses for this report for the following reasons:

- Studies of diagnostic and screening tests were too heterogeneous in terms of outcomes assessed to allow meaningful combination.
- Studies of individual induction agents did not provide sufficient specific information on women in the population of interest. As with diagnostic test studies, there was considerable heterogeneity in terms of outcomes reported.

- We did not identify any significant trials comparing induction to expectant management published subsequent to the most recent Cochrane review (Crowley, 2000). We also did not identify any disagreements with the methods or conclusions of that meta-analysis that were significant enough to justify repeating the analysis.

- Lack of adequate cost data precluded cost-effectiveness analysis.

- Although a decision-analytic model would be an excellent method for exploring the tradeoffs involved in decisionmaking for management of prolonged pregnancy, the considerations discussed above meant that there would be considerable uncertainty surrounding key parameter estimates. While development of such a model even in the setting of widespread uncertainty has considerable value, our past experience with exploratory models in situations where the literature had similar limitations has been that they are of somewhat limited value in further explaining the specific findings of the report.

The approach used by the Cochrane Collaboration differs from ours primarily in the consistent use of meta-analytic techniques to provide summary estimates of the effectiveness and risks of interventions considered. As stated above, we concluded that the state of the literature either could not support meaningful quantitative synthesis relevant to the specific patient population being considered, or that repeating an already well-done meta-analysis (Crowley, 2000) would not be worthwhile. Where relevant Cochrane reviews exist, we have compared their findings and conclusions with our own. Any differences between our findings and Cochrane analyses may represent different inclusion/exclusion criteria, different patient populations considered, or differences in outcomes considered. We have attempted to identify these potential sources of disagreement wherever possible.

Chapter 3. Results

This chapter presents the results of our review, organized around the key questions.

Question 1: What are the test characteristics (reliability, sensitivity, specificity, predictive values) and costs of measures used in the management of prolonged pregnancy to (a) assess risks to the fetus and mother of prolonged pregnancy, and (b) assess the likelihood of a successful induction of labor?

Approach

Assessment of Risks to Fetus and Mother

In Chapter 1, we discussed the evidence for increasing risk of adverse outcomes, especially perinatal death, as gestational age advances beyond 40 weeks. Although this risk is small in absolute terms, the trend towards increasing risk with increasing gestational age is consistent across studies. One approach to preventing these adverse outcomes would be to use testing to identify patients most likely to experience them.

Which antenatal testing strategies lead to improvements in fetal and maternal outcomes? The best way to answer this question is with studies that directly compare one testing strategy with another (or no testing), with the least biased assessment from a randomized control trial, followed by concurrent nonrandomized cohort comparisons, historical cohort comparisons, and cohort studies with variation in testing strategies employed (Evidence Table 1).

However, most of the published literature consists of case series or cohort studies in which there is little or no variation in testing strategies (or variation is not reported). Such studies are less useful but still may contain valuable information concerning the association of test results with fetal and maternal outcomes.

This association can take one of two forms, either prediction of future outcomes (for example, association of antenatal nonstress test [NST] with low Apgar scores or neonatal mortality) or assessment of current status (e.g., measuring abdominal circumference in utero by ultrasound to assess incidence of macrosomia or fetal weight). These studies address the question, “How accurate is the assessment of current fetal status or prediction of future maternal and fetal outcomes offered by antenatal testing?” While evidence that one test is more accurate or has a stronger association with relevant outcomes suggests that it would be more effective, this is by no means definitive. Nevertheless, most of the studies providing data about the predictive value of the tests considered provided 2-by-2 table data (Table 6).

Reliability of Tests

We additionally sought data on the reliability of tests, including interobserver variation, when these were available. If a test result is not reproducible when the test is performed by different examiners, or by the same examiner on different occasions, then the utility of the test is reduced, even if the “average” test characteristics (sensitivity, specificity) imply useful discrimination or prediction.

Correlation of Tests

In certain cases, the association of one test result with another was reported without reference to outcomes.

Results

Assessment of Risks to the Fetus Associated with Uteroplacental Insufficiency

Testing versus no testing. We did not identify any randomized trials in which women with prolonged gestation were randomly assigned to antepartum surveillance or no testing. Of four randomized trials of antepartum cardiotocography versus no surveillance in "high-risk" pregnancies (Brown, Sawers, Parsons, et al., 1982; Flynn, Kelly, Mansfield, et al., 1982; Kidd, Patel, and Smith, 1985; Lumley, Lester, Anderson, et al., 1983)—also the subject of a systematic review by Pattison and McCowan (2001)—only one (Flynn, Kelly, Mansfield, et al., 1982) included patients who were being followed explicitly for prolonged gestation (classified as "suspect postmaturity syndrome" in the paper). In this trial, 100 of 300 subjects were being followed for this indication. All patients received either outpatient ("at intervals of not more than 1 week") or inpatient ("at least twice per week") NSTs. Patients were randomized to two groups: in one, clinicians taking care of the patients knew the results of the NST, while in the other group, NST results were not revealed. Although quantitative data were not reported on this, it appears that the majority of the patients with prolonged gestation received outpatient testing between 41 and 42 weeks, when induction was scheduled.

Although results were not reported separately for women with prolonged gestation, there were no statistically significant differences in stillbirths, neonatal deaths, or other adverse neonatal outcomes between the two groups. However, patients in the group in which caregivers knew the results were significantly more likely to be discharged from the hospital before delivery and significantly more likely to receive outpatient care. There also were nonsignificant trends towards fewer antenatal inpatient days and fewer elective cesarean sections in the group whose caregivers were aware of their results.

In this study (Flynn, Kelly, Mansfield, et al., 1982), a nonreactive NST had 100 percent sensitivity for stillbirths with nonlethal congenital abnormalities and a specificity of 88 percent; positive predictive value was nine percent, and negative predictive value 100 percent. None of the deaths were in the prolonged pregnancy group. Test characteristics for surrogates of fetal compromise were less favorable. For fetal distress in labor, sensitivity was 37 percent, specificity 88 percent, positive predictive value 18 percent, negative predictive value 93 percent. Similar trends were seen for meconium and admission to the neonatal intensive care unit: considerably lower sensitivity than specificity, poor positive predictive value, and good negative predictive value. These findings suggests that the effects on management observed in this trial—consistent trend towards less aggressive observational strategies in the group where the results were revealed to clinicians—reflect clinically appropriate interpretation of the test results. The high negative predictive values are evidence that a normal test does provide reassurance. Unfortunately, the paper does not allow estimation of test characteristics in the specific population of interest for this report, patients with prolonged pregnancy and no other risk factors.

We did identify two retrospective concurrent cohort studies comparing testing and no testing in women with prolonged pregnancy (Bochner, Williams, Castro, et al., 1988; Fleischer, Schulman, Farmakides, et al., 1985). Fleischer, et al., reported a retrospective cohort study comparing 228 women who had weekly NST monitoring beginning at 41 weeks with 30 women who had no antenatal monitoring (Fleischer, Schulman, Farmakides, et al., 1985). Reasons for women not receiving testing were not specified. Despite the small sample size of the no-testing group, the investigators observed significant differences in most of the outcome variables they reported, including low Apgar score (< 7) at 1 and 5 minutes, neonatal intensive care unit (NICU) admission rates, stillbirth rates, and cesarean section for fetal distress. The small sample of women with no monitoring, the retrospective nature of the study design, and the unusually high rates of adverse fetal and maternal outcomes all suggest that the no-testing group in this study may be dissimilar to the NST monitoring group in other ways besides whether an antenatal NST was conducted. This potential confounding probably exaggerates the effectiveness of NST monitoring.

Bochner, et al., described a comparison of large concurrent cohorts of women who underwent antenatal testing with amniotic fluid volume (AFV) and nonstress testing beginning at week 41 or 42 and those with no antenatal testing (Bochner, Williams, Castro, et al., 1988). They found an association with total number of adverse outcomes (testing, 0/512; no testing, 13/1807 [0.7 percent]; $p < 0.05$) and a trend toward higher cesarean section for fetal distress in the no-testing cohort (testing, 14/512 [2.7 percent]; no testing, 60/1807 [3.3 percent]; $p = 0.07$). When the results of testing were compared in the groups beginning testing at 41 weeks ($n = 908$) and those at 42 weeks ($n = 352$), the positive predictive value for a diagnosis of intrapartum fetal distress was significantly higher at 42 weeks (21.1 percent at 42 weeks vs. 11.9 percent at 41 weeks), with a concomitantly lower negative predictive value (98.5 percent at 42 weeks vs. 99.1 percent at 41 weeks). This is consistent with an overall increased risk of adverse outcomes with increasing gestational age, assuming that the sensitivity and specificity of the test are independent of gestational age (more on this below). It is unclear why the no-testing group did not receive testing, since women with “high risk factors” were excluded, and inclusion criteria required that women be seen prior to 20 weeks. Again, the possibility of confounding cannot be ruled out.

In summary, it is difficult to draw conclusions about the effectiveness of antepartum testing compared with no testing in prolonged pregnancy. The only randomized trial comparing testing with no testing is limited by a heterogeneous population (in terms of other risk factors), relatively small numbers of patients with prolonged pregnancy alone, failure to report results separately by indication for testing, and questions about the applicability of the results to current practice (Pattison and McCowan, 2001). The two nonrandomized studies identified suggest an excess risk of adverse outcomes in unmonitored pregnancies, but the failure to characterize the groups studied makes it impossible to rule out other factors as the cause of this excess risk.

Maternal sensation of fetal movement (kick counts). We identified only one study that assessed the association of maternal sensation of fetal movement with postmaturity syndrome, defined as characteristic skin changes (desquamation, leather-like consistency, little subcutaneous fat) and a “long, lean body,” with a ponderal index (weight in grams x 100/length in cubic centimeters) of 2.27 or less (10^{th} percentile or less). Rayburn, et al., tested a group of 147 women at 42 weeks or more gestational age using the NST plus fetal movement charting plus urine estrogen-to-creatinine ratio (Rayburn, Motley, Stempel, et al., 1982). These tests were

performed semi-weekly or weekly. If the NST was reactive (two adequate accelerations of baseline fetal heart rate [FHR] during a 20- to 40-minute period), then it was repeated on the next visit. If the NST was nonreactive, then the test was either repeated or a contraction stress test (CST) was given on the same day. Of the 147 cases studied, 32, or 22 percent, had postmaturity syndrome. However, none of the mothers recording kick counts noted reduced fetal movement (sensitivity, 0/32; specificity, 115/115 [100 percent]). The kick count measure was not useful for predicting postmaturity syndrome, with an undefined positive predictive value and negative predictive value of 78 percent. No studies documenting the reliability of this method (such as correlation between maternal sensation of movement and observed movements on ultrasound) were identified.

In summary, there are no data to suggest that maternal sensation of fetal movement is useful in predicting which infants are affected by postmaturity syndrome. There are no data at all to allow evaluation of maternal sensation of fetal movement as a predictor of other adverse outcomes associated with prolonged gestation.

Nonstress test (NST). We identified one randomized trial enrolling 287 patients comparing the NST alone with a simple biophysical profile (NST plus AFV, supplemented by estimates of fetal weight and placental function) (Arias, 1987). In this trial, 44 of 217 patients had abnormal results on antenatal testing, 14/112 in the NST alone group and 30/105 in the NST + AFV group. There were no significant differences in any outcome, including fetal distress or cesarean section for fetal distress, though slightly more inductions and cesarean sections for fetal distress occurred in the biophysical profile arm. Test characteristics of other components of this combination of tests (ultrasound for fetal weight alone, ultrasound for placental function alone, or ultrasound for AFV alone) were not reported. Sensitivity was similar for NST alone and NST + AFV; however, specificity was higher for NST alone than for NST + AFV. This study was rated positively for 9 of 12 quality assessment items, failing items for sample size and statistical analysis.

Eleven articles provided 40 separate 2-by-2 tables addressing the association of NST with intermediate fetal and maternal outcomes (Arias, 1987; Devoe and Sholl, 1983; Eden, Gergely, Schifrin, et al., 1982; Farmakides, Schulman, Winter, et al., 1988; Fleischer, Schulman, Farmakides, et al., 1985; Phelan, Platt, Yeh, et al., 1984; Ramrekersingh-White, Farkas, Chard, et al., 1993; Small, Phelan, Smith, et al., 1987; Tongsong and Srisomboon, 1993; Weiner, Farmakides, Schulman, et al., 1994; Weiner, Reichler, Zlozover, et al., 1993). The outcomes considered were intermediate in six cases, fetal in 29, and maternal in five cases. The number of specific outcomes is shown in Table 7.

Table 8 shows the sensitivity and specificity, as well as positive and negative predictive values, for each study. For predicting 1-minute Apgar scores < 7, data from five studies (Eden, Gergely, Schifrin, et al., 1982; Fleischer, Schulman, Farmakides, et al., 1985; Phelan, Platt, Yeh, et al., 1984; Small, Phelan, Smith, et al., 1987; Tongsong and Srisomboon, 1993) showed that the sensitivity of NST ranged from 0.12 to 0.41, and specificity ranged from 0.81 to 0.97. For predicting low 5-minute Apgar scores, data from the same five studies and one more (Devoe and Sholl, 1983) showed that the sensitivity of NST ranged from 0 to 0.5, and specificity ranged from 0.80 to 0.95. Two studies used combined endpoints and found that NST was predictive, with sensitivity of 0.08 to 0.33 and specificity of 0.91 to 0.95.

In addition to data on the NST as a whole, two studies reported the predictive value of fetal heart rate monitoring in the context of nonstress testing (Rayburn, Motley, Stempel, et al., 1982; Sherer, Onyeije, Binder, et al., 1998) (Table 9). Neither bradycardia nor tachycardia alone had

high sensitivity or specificity for predicting low Apgar scores, meconium aspiration, or NICU admission. Neither was abnormal heart rate associated significantly with the occurrence of postmaturity syndrome.

In summary, results of these studies suggest that a reactive nonstress test in prolonged pregnancy has good negative predictive value—i.e., adverse outcomes are unlikely to occur in the setting of a reactive nonstress test—but that the positive predictive values are low. Data from the one randomized trial comparing weekly NST beginning beyond 40 weeks to NST and amniotic fluid assessment suggest equivalent outcomes.

Contraction stress test (CST) using oxytocin. Knox, et al., compared the CST using oxytocin with amniocentesis for meconium staining in 187 women at 42 weeks gestation (Knox, Huddleston, and Flowers, 1979). The study was prospective, with women assigned to groups according to the last digit of hospital number. Amniocentesis was obtained on all women at entry into the study, and labor was induced immediately if meconium staining was observed. If no meconium staining was present on initial amniocentesis, then subsequent monitoring was as follows: women in the amniocentesis group received weekly amniocentesis and were induced if meconium staining was present; and women in the CST group received an immediate CST, repeated weekly if normal. Labor was induced in significantly more women in the amniocentesis group than the CST group (11/90 [12 percent] vs. 29/90 [2 percent], respectively; $p < 0.005$). There were no statistically significant differences between testing groups for any outcome, including Apgar score < 7 at 1 minute, Apgar score < 7 at 5 minutes, low birthweight (< 10^{th} percentile), neonatal morbidity, perinatal death, cesarean sections, or abnormal labor (prolonged latent phase, primary dysfunctional labor, secondary arrest of dilatation, or arrest). However, the proportion of babies with Apgar scores less than 7 at 1 and 5 minutes was two-fold higher in the amniocentesis group; the study may have been underpowered to detect this difference.

A single observational study (Devoe and Sholl, 1983) correlated CST results with the clinical outcomes of fetal distress and low Apgar score at 5 minutes (Table 10). Seventy-two of 248 women had labor induced either electively (n = 39) or for abnormal test results (n = 33). Twenty-two women had nonreactive NST followed by positive CST, and 17 women had nonreactive NST but negative CST. The positive predictive value of the CST component of the sequential testing strategy (NST followed by CST if NST is nonreactive) was poor for prediction of low Apgar scores or fetal distress.

In summary, CST is at least equivalent to amniocentesis for meconium staining in terms of outcomes, with significantly fewer inductions; perhaps on the basis of this trial, amniocentesis is no longer used for this indication. In the setting of prolonged pregnancy, CST, when used sequentially for followup of abnormal NST, has good negative predictive value but poor positive predictive value, based on one observational study.

CST using nipple stimulation. We did not identify any studies where nipple stimulation was the sole method for performing contraction stress tests in the management of prolonged pregnancy.

Amniotic fluid measurements. We identified one relevant randomized trial. Alfirevic, et al., compared two ultrasonographic measurements of oligohydramnios, namely amniotic fluid index (AFI) < 7.3 and maximum pool depth (MPD) < 2.1 cm, among 500 women at greater than 40 weeks gestation (Alfirevic, Luckas, Walkinshaw, et al., 1997). Both groups also had NST every

3 days. There were no differences in fetal outcomes between the two strategies; however, abnormal NST was more often an indication for induction in the AFI group than in the MPD group (15 percent vs. 8 percent; $p = 0.04$). The overall rates of induction of labor were not statistically different between groups (87/250 vs. 77/250; $p = 0.39$). There was a trend toward cesarean section for fetal distress being more common in the AFI group than in the MPD group (8 percent vs. 4 percent; $p = 0.09$). One possible explanation for this is a lower threshold for a diagnosis of fetal distress or for performing cesarean section in the presence of nonreassuring fetal heart rate tracings or abnormal antepartum NST results. Since such results were more common in the AFI group, it is not surprising that cesareans for fetal distress also were more common.

In a comparative cohort study, Eden, et al., reported a series of 585 patients managed in one of three ways (based on temporal changes in the protocol used): (1) weekly NST with CST for nonreactive NST (from November 1, 1978 through August 31, 1979); (2) semi-weekly NST with biophysical profile for nonreactive NST (from September 1, 1979 through December 31, 1980); or (3) semi-weekly NST with biophysical profile for nonreactive NST, plus weekly AFV measurement (from January 1, 1981 through August 31, 1981) (Eden, Gergely, Schifrin, et al., 1982). The groups employing the biophysical profile had lower incidences of low Apgar score at 5 minutes, meconium aspiration, stillbirth, fetal distress requiring intervention (persistent abnormal FHR patterns), and morbidity (defined as presence of any of following: fetal distress requiring intervention, 5-minute Apgar score < 7, neonatal resuscitation, postmaturity syndrome, or meconium aspiration). However, the rate of cesarean sections was significantly higher in the groups using the biophysical profile than in the group using NST + CST alone (NST + CST, 11.5 percent; NST + biophysical profile, 29.9 percent; NST + AFV + biophysical profile, 29.4 percent; 1 vs. 2, $p < 0.05$; 1 vs. 3, $p < 0.05$). This suggests that tests using the biophysical profile may be more sensitive at identifying fetuses at risk, but that subsequent induction resulted in higher cesarean section rates. Alternatively, as discussed above, physician thresholds for performing cesarean section may be quite different based on knowledge of antepartum test results. Despite the higher rates of cesarean section, the incidence of fetal distress requiring intervention was substantially lower in the groups using biophysical profile testing in addition to NST (NST + CST, 21.8 percent; NST + biophysical profile, 4.5 percent; NST + AFV + biophysical profile, 5.5 percent; 1 vs. 2, $p < 0.05$; 1 vs. 3, $p < 0.05$).

Tongsong and Srisomboon (1993) performed NST and AFV in 242 women at 42 weeks or more in gestational age. AFV was more accurate than NST in predicting intrapartum fetal distress ($p < 0.05$) (AFV: sensitivity, 73 percent; specificity, 91 percent; positive predictive value, 27 percent; negative predictive value, 99 percent; NST: sensitivity, 64 percent; specificity, 82 percent; positive predictive value, 14 percent; negative predictive value, 98 percent). Given that the definition of intrapartum fetal distress included moderate to severe variable decelerations, which would be more likely in a setting of oligohydramnios, which in turn would be more likely to be detected with ultrasound, these results are not surprising.

Table 11 summarizes sensitivity, specificity, and positive and negative predictive values for predicting reported perinatal and maternal outcomes, using amniotic fluid measurement with various criteria for abnormality. In general, specificity is markedly better than sensitivity, while negative predictive value is better than positive predictive value, as was also the case with NST and CST.

Abdominal palpation. As part of an investigation of the value of ultrasound evaluation of amniotic fluid volume in predicting adverse outcomes, Crowley, et al., also evaluated the performance of clinical assessment of AFV by abdominal palpation. This technique had a false positive rate of 25 percent and a false negative rate of 43 percent for predicting "significant meconium staining or absent amniotic fluid" at the time of amniotomy (Crowley, O'Herlihy, and Boylan, 1984).

Simple biophysical profile. Table 12 describes the individual components of the various biophysical profiles employed in the studies included in this report. One randomized trial and four noncomparative studies provide data on a simple biophysical profile (NST plus measurement of amniotic fluid volume). The randomized trial compared a simple biophysical profile (NST + maximum pool depth [MPD]) with a complex biophysical profile consisting of NST, amniotic fluid index (AFI), fetal breathing movements, fetal tone, and fetal gross body measurements for antenatal monitoring (Alfirevic and Walkinshaw, 1995). There were more abnormal test results with the complex biophysical profile (47 percent vs. 21 percent; $p = 0.0013$), more inductions of labor (60 percent vs. 41 percent; $p = 0.04$), and more inductions associated with abnormal testing (39 percent vs. 15 percent; $p = 0.002$). There were no significant differences in clinical fetal or maternal outcomes. Cesarean section rates were nonsignificantly higher in the complex monitoring group (18 percent vs. 10 percent; $p = 0.22$).

Four studies described the accuracy of simple biophysical profiles for predicting a variety of outcomes (Arias, 1987; Bochner, Medearis, Ross, et al., 1987; Bochner, Williams, Castro, et al., 1988; Brar, Horenstein, Medearis, et al., 1989) (Table 13). Although Bochner, et al. (1987) reported high values for sensitivity and specificity of the simple biophysical profile for predicting low Apgar scores at 5 minutes and cesarean section for fetal distress, the confidence intervals around those estimates were wide because the 2-by-2 tables were based on a relatively small subset ($n = 62$) of the study's 845 patients. The other studies show relatively poor sensitivity and specificity.

Table 13 summarizes the results of studies of simple biophysical profiles. Again, in general, specificity for the various outcomes is better than sensitivity, while negative predictive value is consistently higher than positive predictive value.

Complex biophysical profile score. The randomized trial of Alfirevic and Walkinshaw (1995) comparing simple with complex biophysical profiles is discussed above. Three other studies reported data on the performance of a complex biophysical score (Table 14). Since the definition of "complex" varied between studies, the items used to calculate the scores in individual studies are shown in Table 12.

Arabin, Snyjders, Mohnhaupt, et al. (1993) compared the predictive ability of a biophysical profile consisting of NST, amniotic fluid assessment, fetal tone, fetal movements, and fetal breathing to a novel fetal assessment score consisting of five components: FHR pattern, uterine artery resistance by Doppler ultrasound, carotid artery resistance index by Doppler ultrasound, fetal tone (movements) by ultrasound, and fetal reflexes (magnitude and speed of movements) by ultrasound. In receiver operating characteristic (ROC) analysis, the fetal assessment score provided better prediction of fetal distress and low Apgar score at 1 minute than did the biophysical profile ($p < 0.001$) but not better prediction of low umbilical artery pH. Qualitatively, the difference was greatest for prediction of fetal distress, with less difference noted for prediction of low Apgar scores and none for prediction of low pH. This suggests that

the fetal prediction score is better at discriminating results that correlate directly with its component tests (such as fetal distress defined by abnormal fetal heart rate patterns) than at true physiological measures of fetal compromise. One possible explanation for this could be interpretation of intrapartum fetal monitoring based on prior knowledge of antepartum test results.

Hann, et al., reported the results of biophysical profile monitoring in 131 women at 41 completed weeks gestation (Hann, McArdle, and Sachs, 1987). Positive predictive values for "poor neonatal outcome" (neonatal distress requiring admission to the neonatal intensive care unit, endotracheal intubation, use of positive pressure ventilation for more than 6 hours, and/or persistent fetal circulation) for the composite biophysical profile at a threshold of ≤ 6 was 14 percent; for individual components, positive predictive values were as follows: AFV, 17 percent; placental grading, 4 percent; fetal breathing movements, 5 percent; fetal tone/movements, 40 percent; and nonreactive NST, 14 percent. Negative predictive value for the composite biophysical profile was 94 percent; for individual components: AFV, 95 percent; placental grading, 91 percent; fetal breathing movements, 94 percent; fetal tone/movements, 95 percent; and reactive NST, 94 percent.

Gilson, O'Brien, Vera, et al. (1988) describe the association between twice weekly biophysical profile monitoring and low Apgar scores, fetal distress, and cesarean section for fetal distress among 178 women at greater than 42 weeks gestation. At the cut-point used (a score of 8), the test showed poor sensitivity across all outcomes, ranging from 0.08 to 0.27.

Table 14 summarizes the test characteristics reported in these studies. Again, specificity is generally better than sensitivity, while negative predictive value is consistently much higher than positive predictive value.

Doppler measurements of umbilical blood flow. Two studies reported data on the predictive value of Doppler measurements of umbilical artery blood flow (Battaglia, Larocca, Lanzani, et al., 1991; Farmakides, Schulman, Winter, et al., 1988) (Table 15). Battaglia, et al., evaluated Doppler velocimetry of umbilical artery used as screening test for predictive value in a case series (Battaglia, Larocca, Lanzani, et al., 1991). This was performed as a battery of tests including NST; amnioscopy; AFV; Doppler velocimetry of the uterine, umbilical, descending thoracic aorta, renal, and middle cerebral arteries; and a series of maternal blood measurements, including hPL, estriol, hematocrit, platelets, mean platelet volume, and uric acid. The criteria for decisionmaking about induction and delivery were not described. Doppler velocimetry was strongly associated with adverse outcomes, including "poor condition" (both 1- and 5-minute Apgar scores < 7 or infant admitted to NICU for asphyxia and/or meconium aspiration syndrome), oligohydramnios (largest pocket < 2 cm), meconium staining, and cesarean sections for fetal distress. Of note, 4 of 16 of these infants had birthweights greater than 4,000 grams; it is unclear to what extent these infants, who presumably had normal uteroplacental function, affected the results.

Farmakides, et al., reported on 140 high-risk pregnancies (33 percent were postdate) that were followed with NST and Doppler velocimetry (Farmakides, Schulman, Winter, et al., 1988). "Most" of the cases of fetal distress and cesareans for fetal distress came from the postdate subgroup. Nonreactive NST was significantly more sensitive at predicting cesarean section for fetal distress than Doppler. Since management decisions were based on NST results, this again raises the possibility of biased decisionmaking based on prior knowledge of antepartum test results.

Table 15 summarizes the results of these studies of Doppler. Again, negative predictive value is consistently higher than positive predictive value, although sensitivity appears to be improved relative to specificity compared with the other tests reviewed in this report.

Summary of tests to evaluate risks to the fetus associated with uteroplacental insufficiency. There are no randomized trials comparing antepartum testing by any method to no testing in women with prolonged pregnancy only. Data from one relatively large retrospective cohort (Bochner, Williams, Castro, et al., 1988) suggest an increased risk of adverse outcomes to the fetus, although confounding cannot be eliminated as a possibility for this observed association. Evidence from large registries shows consistently elevated risks of antepartum stillbirth with increasing gestational age, even in health systems where testing is available (see the section on "Risk of Perinatal Mortality" in chapter 1). Given this elevated risk, it is highly unlikely that a randomized trial of testing versus no testing could be performed in the United States without, at the least, extreme difficulty with recruitment. The low absolute risk of stillbirth makes sample size requirements prohibitive as well. For example, the estimated perinatal mortality at 41 weeks in terms of deaths per 1,000 ongoing pregnancies is approximately 1.2. A randomized trial would need over 40,000 women in each arm to determine a two-fold difference in risk of stillbirth between two competing methods of antepartum surveillance.

Because of the numerous methodological issues involved in evaluating specific antepartum tests (see discussion below), we are unable to conclude that any test or combination of tests is clearly superior to another. Only one randomized trial directly compared a more complex test with a simpler test (Alfirevic and Walkinshaw, 1995); this trial showed that the more complex test resulted in more interventions with no difference in outcomes. As with most tests, there appear to be consistent tradeoffs between sensitivity and specificity–tests that are more sensitive are likely to be less specific. We did not identify published data on inter- or intraobserver variability of these tests in the specific context of monitoring prolonged pregnancy or on the medical and nonmedical costs associated with specific tests and testing regimens.

We did find that, qualitatively, specificity for most tests was considerably better than sensitivity, while negative predictive value also was considerably better than positive predictive value. This means that women with "normal" test results are highly unlikely to experience the adverse outcomes used to determine a true "positive" test result. The high specificities reported may reflect biases in study design–when outcomes are either directly related to test results (such as nonreassuring fetal heart rate tracings after abnormal antepartum NST) or likely to be influenced by knowledge about the test results (such as cesarean section for fetal distress), specificity is likely to be relatively high.

This pattern of high negative predictive value in the setting of relatively low sensitivities has interesting implications for future management strategies. By Bayes' Theorem, positive predictive value can be expressed as:

True Positives/(True Positives + False Positives), or
[(Prevalence)*(Sensitivity)] /{[(Prevalence)*(Sensitivity)] + [(1-Prevalence)*(1-Specificity)]},
while negative predictive value is expressed as:

True Negatives/(True Negative + False Negatives), or
[(1-Prevalence)*(Specificity)] /{[(1-Prevalence)*(Specificity)] + [(Prevalence)*(1-Sensitivity)]}.

In practice, this means that increasing test sensitivity results in a higher negative predictive value, since the false negative rate decreases. Increasing test specificity results in a higher

positive predictive value, since false positives decrease. Given the consistent pattern observed for all of the reviewed antepartum tests that specificity is higher than sensitivity, one would expect that positive predictive value would be higher than negative predictive value. The fact that the pattern is consistently the opposite suggests that it is the relatively low prior probability of adverse outcomes, the "prevalence" in the equations above, that drives the predictive values.

If this is the case, then the following points need to be considered:

- The main purpose of antepartum testing is primarily to avoid unexplained stillbirths and secondarily to avoid perinatal morbidity. In order to accomplish these things, tests with high negative predictive values are needed. One way to achieve this would be to improve the sensitivity of currently used antepartum testing technologies. Since it is unlikely that sensitivity can be increased without a subsequent decrease in specificity, this means that the positive predictive value of these tests will decrease further.

- If, as the reviewed studies suggest, the probability of adverse outcomes is currently what determines predictive values, then this means that the positive predictive value of antepartum testing will improve and the negative predictive value decline as gestational age increases, since the risk of stillbirth and other adverse events increases with gestational age. This proposition is dependent on the assumptions that (1) sensitivity and specificity are independent of gestational age, and (2) the outcomes reported in these studies are reasonable surrogates for stillbirth risk. This proposition is consistent with the data reported by Bochner, Williams, Castro, et al. (1988), according to which the positive predictive value for all adverse outcomes was better when testing began at 42 weeks (21.1 percent vs. 11.9 percent when testing began at 41 weeks), but the negative predictive value was worse (98.5 percent at 42 weeks vs. 99.1 percent at 41 weeks).

- Assuming that induction of labor does not carry increased perinatal risks compared with spontaneous labor, planned induction of labor at a given gestational age will always result in fewer expected adverse perinatal outcomes compared with testing strategies, since the negative predictive value of the tests will continue to decline as gestational age advances. At earlier gestational ages, where the risk is very low, the number of patients required to demonstrate this would be quite large.

These implications will be discussed further in the context of the trials of induction versus testing (Question 2).

Assessment of Risks to the Fetus and Mother Associated with Fetal Macrosomia

Because both mother and infant are at risk of injury secondary to macrosomia, various methods for estimating fetal weight have been evaluated. Macrosomia is usually defined as a newborn weight of greater than 4,000 grams or 4,500 grams; the clinical significance of birthweights between 4,000 and 4,500 grams is unclear, since risk of shoulder dystocia is greatest for infants over 4,500 grams (ACOG, 2000).

Clinical exam. Chauhan, et al., compared estimates of fetal weight by clinicians using Leopold maneuvers in early labor, sonographic measurements obtained by the same clinicians, and actual birthweight (Chauhan, Sullivan, Magann, et al., 1994). Clinical estimation was significantly more accurate than ultrasound estimation as measured by mean absolute error compared with actual weight (clinical, 322 ± 253 g; sonographic, 547 ± 425 g; $p < 0.001$), mean percentage absolute error (clinical, 8.9 ± 7.1 g/kg; sonographic, 14.8 ± 11.0 g/kg; $p < 0.001$), and percentage of estimates within 10 percent of actual birthweight (clinical, 65.4 percent; sonographic, 42.8 percent; $p < 0.005$).

The same group also compared maternal estimations by women with prior childbearing experience with clinical estimation (Chauhan, Sullivan, Lutton, et al., 1995). There were no significant differences in the accuracy of maternal estimates compared with clinical estimates.

Ultrasound. Chauhan, et al. (Chauhan, Sullivan, Magann, et al., 1994) found that clinical estimation was more accurate than ultrasonographic estimation by the same clinician (see above). Ultrasound was slightly more sensitive at predicting birthweight greater than 4,000 grams (55 percent vs. 50 percent, based on 20 cases).

Chervenak, et al., compared 317 women followed for prolonged pregnancy with twice weekly NST and AFT with100 control patients delivered between 38 and 40 weeks (Chervenak, Divon, Hirsch, et al., 1989). Fetal weights were also obtained, although it is unclear how often these measurements were performed. Overall incidence of birthweight greater than 4,000 grams was significantly higher in postdate patients (24 percent vs. 4 percent; $p < 0.05$), and cesarean section rates for arrest or protraction disorders were significantly higher when infants weighed more than 4,000 grams (22 percent vs. 10 percent; $p < 0.01$). Sensitivity of ultrasound for predicting birthweight greater than 4,000 grams was 61 percent, specificity 91 percent, positive predictive value 70 percent, and negative predictive value 87 percent. Morbidity associated with macrosomia was not reported. It is unclear to what extent clinicians managing the patients had access to the ultrasound reports. Since clinicians might have a lower threshold for diagnosing an arrest or protraction disorder in the setting of suspected macrosomia, this would result in a bias in favor of improved positive predictive value for ultrasound.

Gilby, et al., constructed ROC curves for the performance of two abdominal circumference cut-points (35 cm and 38 cm) for predicting macrosomia at two thresholds, 4,000 grams and 4,500 grams, from a series of 1,996 subjects who had ultrasounds within 7 days of delivery (Gilby, Williams, and Spellacy, 2000). At a cut-point of 35 cm, sensitivity for prediction of birthweight of 4,500 grams was 98.5 percent, specificity 64.6 percent, positive predictive value 9.1 percent, and negative predictive value 99.9 percent. At a cut-point of 38 cm, sensitivity was 53.6 percent, specificity 96.8 percent, positive predictive value 37.3 percent, and negative predictive value 98.3 percent. Morbidity associated with macrosomia was not reported. Whether these predictive values would be applicable in a different population is unclear.

O'Reilly-Green and Divon (1997) constructed ROC curves for ultrasonographic estimates of fetal weight, with an adjustment of 12.7 grams added to the estimated fetal weight (EFW) for each day elapsed between sonographic measurements and delivery. Areas under the ROC curve for prediction of birthweight greater than 4,000 grams were 0.85 and 0.93 to 0.95 for prediction of birthweight greater than 4,500 grams, indicating good discriminative ability. Relatively small relative increments in EFW had large impacts on sensitivity and specificity: for prediction of actual birthweight of greater than 4,000 grams, an EFW of 3,711 grams had a sensitivity of 85 percent and specificity of 72 percent, while an EFW of 4,000 grams had a sensitivity of 56

percent and a specificity of 91 percent. For prediction of birthweight greater than 4,500 grams, an EFW of 4,192 grams had sensitivity of 83 percent and specificity of 92 percent, while an EFW of 4,500 grams had a sensitivity of 22 percent and a specificity of 99 percent. Again, no correlation with outcomes associated with fetal macrosomia were reported.

Test performance characteristics for studies reporting association between estimated fetal weight and macrosomia are shown in Table 16.

Summary: Tests for predicting fetal macrosomia. There is a clear tradeoff between sensitivity and specificity of markers for estimating fetal weight. The definition of macrosomia also plays a role. In studies in women with prolonged pregnancy, sensitivities for detection of birthweight greater than 4,000 grams range from 56-89 percent, with specificities of 72-93 percent; positive predictive values at this threshold range from 49-93 percent, with negative predictive values of 87-94 percent. At a threshold of 4,500 grams, sensitivity ranges from 14-99 percent and specificity from 65-99 percent, with positive predictive values of 9-44 percent and negative predictive values of 96-100 percent. Positive predictive value at the more clinically significant 4,500 gram threshold is worse than at 4,000 grams (not surprisingly, since the probability of a weight greater than 4,500 grams is much lower than for 4,000 grams). However, translation of even this diagnostic test accuracy into clinical strategies that significantly reduce injury risk to either mother or infant at an acceptable cost in terms of iatrogenic complications or resource use is difficult.

Prior suspicion of fetal macrosomia does not appear to result in improved outcomes for either mother or infant. Weeks, et al., reported a retrospective series of 504 infants with birthweight greater or equal to 4,200 grams (Weeks, Pitman, and Spinnato, 1995). In 102 patients, macrosomia was suspected, while it was not in the remaining 402. Cesarean delivery rates were significantly higher in the suspected group (52 percent) compared with the unsuspected group (30 percent), a difference attributable to a higher rate of labor induction and failed induction. Among patients undergoing vaginal delivery, shoulder dystocia occurred in 24.5 percent of the predicted group and 16.7 percent in the not predicted group, a difference that was not statistically significant (which may be due to lack of power).

Even better evidence of a lack of benefit comes from a trial in which women at 38 weeks or more with estimated birthweights between 4,000 and 4,500 grams based on ultrasound were randomized to either immediate induction or expectant management. There were no statistically significant differences in cesarean delivery rate, instrumental delivery rate, or incidence of shoulder dystocia between the two groups (Gonen, Rosen, Dolfin, et al., 1997). There were trends toward higher instrumental delivery rates in induced nulliparous women (26.2 percent vs. 15 percent in expectantly managed nulliparous women) and higher cesarean section rates in expectantly managed multiparous women (16.2 percent vs. 10.9 percent in induced multiparous women). Other maternal outcomes, such as perineal or vaginal trauma, were not reported. The study was underpowered to detect differences in neonatal morbidity; overall rates were low (9/134 in the induction group and 11/139 in the expectant group), with six or fewer cases of any single type of morbidity (cephalohematoma, with nine cases, was most common).

Rouse, Owen, Goldenberg, et al., (1996) estimated based on available data that a policy of elective cesarean section for an estimated fetal weight of 4,500 grams or more would result in 3,695 cesarean deliveries at a cost of over $8 million to prevent one permanent brachial plexus injury.

In summary, methods for detection of macrosomia defined as birthweight greater than 4,500 grams are imprecise. There is evidence that clinical measurements, including multiparous patients' own estimates, are as accurate as ultrasound. Available data suggest that there is no benefit to mother or infant from induction of labor for suspected macrosomia (when defined as estimated weights between 4,000 and 4,500 grams). While an estimate of fetal weight in theory may have some benefit in management of labor (such as avoidance of operative vaginal deliveries in settings where shoulder dystocia risk is higher), available observational data suggest that suspicion of macrosomia prior to labor does not improve outcomes. There is no evidence that ultrasonographic measurement of fetal weight to detect macrosomia in the setting of prolonged pregnancy improves maternal or neonatal outcomes.

Assessment of the Likelihood of Successful Induction

Cervical examination (Bishop score). The Bishop score was first reported in 1964 as a predictor of the likelihood of a successful induction (Bishop, 1964). The score is based on five components: cervical dilation, cervical effacement, cervical consistency, cervical position, and fetal station (Table 17).

In Bishop's original report (Bishop, 1964), induction was successful in 100 percent of cases (no denominator given) when the Bishop score was greater than 9. Data for lower scores were not given, and notably, all inductions were apparently in multiparous patients, since "[o]wing to the unpredictability of the duration of labor in the nullipara, even in the presence of apparently favorable circumstances, induction of labor brings little advantage for either obstetrician or patient." There was a statistically significant negative correlation between score and interval from examination to spontaneous delivery, but confidence intervals were quite wide (quantitative data were not provided, only a graphic representation).

Three studies provided limited data on the predictive value of Bishop scores (Harris, Huddleston, Sutliff, et al., 1983; Mouw, Egberts, Kragt, et al., 1998; Witter and Weitz, 1989). Harris, et al., reported that dilatation, effacement, and station were more predictive of interval between examination and spontaneous delivery in prolonged pregnancy than consistency and position (Harris, Huddleston, Sutliff, et al., 1983). Witter and Weitz (1989) found that Bishop scores at baseline in women induced at 42 weeks were statistically significantly lower in women who underwent cesarean delivery than in those with vaginal delivery, but that the absolute difference was small; significant overlap made the test a poor discriminator of successful induction (Table 18). Mouw, et al., reported that a Bishop score greater than 5 at 41 weeks had sensitivity 0.67 (95 percent CI, 0.48 to 0.82) and specificity 0.77 (95 percent CI, 0.54 to 0.92) for predicting birth within 3 days; however, only 74 percent of patients in this study had Bishop scores recorded (Mouw, Egberts, Kragt, et al., 1998).

The relatively poor discrimination of the Bishop score in predicting either labor or subsequent successful induction in prolonged pregnancy is magnified by the inherent unreliability of many of its component measures. Significant interobserver variability has been reported in measurement of cervical effacement (Goldberg, Newman, and Rust, 1997; Holcomb and Smeltzer, 1991). Furthermore, significant intra- and interobserver variability has been described for assessment of cervical dilatation (Phelps, Higby, Smyth, et al., 1995; Tuffnell, Bryce, Johnson, et al., 1989)

Fibronectin. Three studies were identified that evaluated the possible use of fetal fibronectin (fFN) obtained from cervicovaginal secretions, a sensitive marker for impending labor, in the management of prolonged pregnancies (Table 19). Tam, et al., measured fetal fibronectin in 58 women at term or beyond, scheduled for induction with PGE_2 suppositories (Tam, Tai, and Rogers, 1999). Thirty women were negative and 28 positive for fibronectin prior to the placement of the suppositories. There was a trend towards a higher gestational age in fibronectin-positive patients (median 294 days, range 280-294, compared with a median of 281 days, range 272-294, in negative patients). Median interval from induction to delivery was significantly lower in fibronectin-positive patients (760 minutes vs. 1,285 minutes). Fibronectin positivity was a reasonable predictor of vaginal delivery (sensitivity 36 percent; specificity 79 percent; positive predictive value 84 percent; negative predictive value 28 percent). Results in this study were not stratified by gestational age or by indication for induction.

Mouw, et al., measured fetal fibronectin at 41 weeks (Mouw, Egberts, Kragt, et al., 1998). A positive fFN test (≥ 50 ng/ml) had sensitivity of 0.71 (95 percent CI, 0.58 to 0.86) and specificity of 0.64 (95 percent CI, 0.48 to 0.78) for predicting birth within 3 days. The change from negative to positive fFN values often occurred between 1 and 4 days before birth in women with a spontaneous onset of labor. The mean interval between positive test and birth was 2.5 ± 2.5 days (range, 0-11).

Imai and colleagues measured vaginal fFN and a panel of cytokines (interleukin 1-beta, interleukin-6, interleukin-8, and tumor necrosis factor alpha) weekly in 122 women from 36 through 42 weeks (Imai, Tani, Saito, et al., 2001). Vaginal fFN was inversely correlated with sampling to delivery interval (r = -0.40). At a threshold of > 50 ng/ml, fFN had a sensitivity of 90 percent, a specificity of 50 percent, a positive predictive value of 75 percent, and a negative predictive value of 75 percent for predicting delivery within 7 days. Interleukin 1-beta was the only cytokine with reasonable performance, but it was less able to discriminate than fFN (sensitivity 55 percent, specificity 76 percent). Results were not stratified by parity or gestational age.

Summary: Tests for assessing the likelihood of successful induction. The Bishop score has a long history in obstetric decisionmaking. Clearly, clinically detectable changes in the cervix take place prior to the onset of labor, and the likelihood of a successful induction should be greater the closer a given patient is to spontaneous labor. However, the documented substantial inter- and intraobserver variability in the components of the Bishop score suggest that its ability to discriminate between women likely to have a successful induction of labor and those unlikely to have a successful induction may be relatively poor. Certainly, given this inherent variability and the discrete nature of its components, changes in the global Bishop score are less than satisfactory primary outcomes for studies of induction or cervical ripening agents.
Data on the clinical utility of fetal fibronectin as a decisionmaking tool in managing prolonged pregnancy are insufficient to draw conclusions. Fetal fibronectin may have potential as a tool for helping to identify women likely to deliver spontaneously within the next 7 days, which in turn may help guide decisionmaking about antepartum testing versus induction.

Methodological Issues

Study Design

- Choice of appropriate outcome measures: Many of the most important outcome measures, especially stillbirth, are so rare that studies using these outcomes are almost impossible to perform. Surrogate markers therefore are not inappropriate, but their clinical relevance is not always clear. For example, although meconium aspiration is a significant adverse outcome with potential for long-term negative sequelae, the presence of meconium-stained amniotic fluid alone is not. Intrapartum abnormal fetal heart rate tracings themselves are subject to significant observer variability (Ayres-de-Campos, Bernardes, Costa-Pereira, et al., 1999; Bernardes, Costa-Pereira, Ayres-de-Campos, et al., 1997; Donker, van Geijn, and Hasman, 1993; Lidegaard, Bottcher, and Weber, 1992), and interpretation may be influenced by prior knowledge of antepartum test results, making fetal heart rate patterns, or cesarean section decisions based on these patterns, less than ideal as surrogate markers of fetal compromise.

- Bias: Many of the studies reviewed either did not state whether clinicians managing patients were aware of test results or definitely stated that these results were available. Since knowledge of these results could affect both interpretation of outcomes (as discussed above) or thresholds for decisionmaking (e.g., greater reluctance to use oxytocin to augment labor if prior antepartum testing was abnormal, or a lower cesarean section threshold for arrest of dilatation or descent if macrosomia were suspected), the ability of tests to predict these outcomes could be falsely elevated.

- Resource use: Data on the medical and nonmedical costs of any of the tests reviewed are lacking.

Statistical Issues

- Inappropriate summary measures and tests: Many studies used means or t-tests for variables such as Bishop scores, Apgar scores, or parity, where values other than integers are meaningless.

- Sample size: Few studies discussed sample size issues.

- Failure to account for variability: No study attempted to account for the effects of observer variation on the precision of estimates. For tests where quantitative values are used to establish a threshold for normal and abnormal, this variability will have implications for the precision of sensitivity and specificity.

Summary

- The risk of antepartum stillbirth clearly increases with increasing gestational age. Although definitive evidence that antepartum testing at some point after 40 weeks reduces perinatal mortality is not available, there are some data consistent with an increased risk of adverse

outcomes in women who do not get tested (Bochner, Williams, Castro, et al., 1988; Fleischer, Schulman, Farmakides, et al., 1985). The most appropriate time to begin antepartum testing in otherwise low-risk women is unclear. An excellent decision analysis of antepartum testing in high-risk women prior to 40 weeks illustrated that the tradeoffs are between the risk of stillbirth, the risk of neonatal death, and the sensitivity and specificity of the test (Rouse, Owen, Goldenberg, et al., 1996). Since the risk of neonatal death in an otherwise uncomplicated pregnancy at term is quite low, the main issues are the stillbirth risk and test characteristics. Unfortunately, our review does not allow precise estimation of the test characteristics of any of these tests in detecting infants at greatest risk for stillbirth in otherwise uncomplicated pregnancies after term.

- As the sensitivity of antepartum testing for predicting surrogate markers of fetal compromise increases, specificity decreases. Testing strategies involving a combination of fetal heart rate monitoring and ultrasonographic measurement of amniotic fluid volume appear to have the highest levels of sensitivity; however, methodological issues and variability in specific tests and testing strategies prohibit definitive conclusions about which test or combination of tests has the best performance.

- Qualitatively, we found that specificity was much higher than sensitivity for most of the outcomes measured, but negative predictive values were much higher than positive predictive values, suggesting that outcome probability is currently the most important determinant of test performance. This in turn implies that the negative predictive value will decrease as gestational age advances, and rates of adverse outcomes due to false negative test results will increase, if sensitivity and specificity of antepartum tests are independent of gestational age. Identifying the most appropriate time to begin testing (or to consider induction) is ultimately dependent on identifying threshold risks of adverse outcomes when weighed against the risks and costs of intervention. We did not identify any data that would allow estimation of that threshold risk.

- Low positive predictive values mean that intervention rates will be relatively high. The degree to which individual women, or society, are willing to trade off risk of adverse fetal outcomes due to prolonged pregnancy, versus the potential for iatrogenic adverse outcomes associated with interventions, is unclear. How variability in the value women place on the nature of the process of labor and delivery (minimal intervention vs. use of the full range of available obstetric, anesthetic, and pediatric technologies) factors into decisionmaking is also unclear.

- Clinical assessment is equivalent to ultrasound in predicting macrosomia. However, there is no evidence that prior knowledge of estimated fetal weight improves outcomes for either infant or mother.

- Clinical examination of the cervix may help predict successful induction. However, individual components of the examination exhibit substantial inter- and intraobserver variability.

- Published data do not allow estimation of the cost-effectiveness of tests of fetal wellbeing.

Question 2: What is the direct evidence comparing the benefits, risks, and costs of planned induction versus expectant management at various gestational ages?

Approach

As with all of the questions addressed in this report, the issue of the appropriate gestational age to consider " postdate" or "postterm" was difficult to resolve. After extensive discussion with the project's advisory panel, a consensus was reached that we would include any articles where the proposed benefit of the planned induction was reduction in maternal or fetal risk associated with prolonged pregnancy, even at 40 weeks gestation. Active interventions performed prior to or shortly after term (such as nipple stimulation or membrane sweeping) that are designed to decrease the proportion of women who go beyond 41 or 42 weeks are discussed under Question 3, below.

Up to this point in the report, we have:

- Found evidence from observational studies of an increasing risk of adverse perinatal events as gestational age advances beyond term. Although the precise degree of this risk is unclear and may be affected by confounding, the pattern is quite consistent.

- Found in our review of antepartum tests of fetal well being in prolonged pregnancy that the sensitivity of such tests was much lower than the specificity, while the negative predictive value was much higher than the positive predictive value.

- Discussed the fact that these two findings, when taken together, suggest that the negative predictive value of antepartum testing will decrease as gestational age advances.

If negative predictive value does decrease with advancing gestational age, then elective induction has the potential to improve outcomes by preventing adverse perinatal outcomes due to false negative test results. Whether this is the case, and whether elective induction is associated with an excess of other adverse maternal outcomes compared with expectant management and testing, is the focus of this section of the report.

Throughout this section, we use the term "expectant management," as defined by the authors of the studies reviewed, to refer to some form of ongoing assessment of fetal well being, with induction of labor based on the results of testing or upon reaching a specified gestational age in accordance with a predefined set of guidelines. As stated above, we did not identify any randomized trials that provided data on the specific population of interest where no intervention (induction or testing) was performed.

As with studies of testing, the outcomes assessed in these trials were quite variable. All studies reported on perinatal mortality and cesarean section rates, in some cases stratified by indication for induction (elective or based on abnormal test results). Additional markers of perinatal or maternal morbidity—including Apgar scores at 1 and 5 minutes, umbilical arterial pH, the presence of meconium-stained amniotic fluid, abnormal fetal heart rate tracings during labor, instrumental deliveries, diagnosis of meconium aspiration, and admissions to neonatal intensive care units—were inconsistently reported.

None of the included trials was able to blind physicians, midwives, and nurses to the allocated intervention or to the results of antepartum testing. Because of this, outcomes that are dependent on interpretation of fetal monitoring (such as the proportion of cesarean sections performed for fetal distress, or the overall incidence of abnormal fetal heart rate tracings) are unreliable. A diagnosis of fetal distress may be more likely in the setting of an induction performed in the expectant management arm after abnormal antepartum monitoring. Even with a normal intrapartum tracing, thresholds for performing cesarean section or operative vaginal delivery in the setting of prolonged second or third stages of labor might be different if the provider is aware of previous abnormal antepartum tests. Because of these difficulties, we focus on the overall cesarean section rate and neonatal outcomes less susceptible to bias, such as the Apgar score, pH, and admissions to the neonatal intensive care unit. Even these immediate outcomes do not provide information on the impact of maternal interventions on longer-term health outcomes of these children.

Results

Trials Identified

The literature search identified 17 relevant publications reporting on 15 separate trials (see Evidence Table 2). In two cases, initial trial reports were followed by publications describing further analyses conducted on the same populations: Pearce and Cardozo (1988) reported the results of supplementary analyses conducted on the population first described by Cardozo, Fysh, and Pearce (1986), and Goeree, Hannah, and Hewson (1995) reported the results of a cost-effectiveness analysis of data collected during the Canadian Multicenter Post-term Pregnancy Trial (Hannah, Hannah, Hellmann, et al., 1992).

The included trials were published between 1983 and 1997. The number of subjects in each trial was fairly small, except for the Canadian trial (Hannah, Hannah, Hellmann, et al., 1992). The overall median number of subjects was 200, ranging from 22 (Martin, Sessums, Howard, et al., 1989) to 3,418 (Hannah, Hannah, Hellmann, et al., 1992).

Benefits

Effects on perinatal mortality. The included studies suggest that induction results in fewer perinatal deaths than does expectant management. Table 20 summarizes perinatal deaths not due to congenital abnormalities in the two management groups. There were a total of seven deaths in the monitoring group compared with no deaths in the induction group.

A meta-analysis performed as part of a recent Cochrane review (Crowley, 2000) showed that this reduction in perinatal mortality with induction is significant only at 41 weeks or later (summary odds ratio [OR], 0.13; 95 percent confidence interval [CI], 0.01 to 2.07 before 41 weeks vs. summary OR, 0.23; 95 percent CI, 0.06 to 0.90 at 41 weeks or later).

Effects on perinatal morbidity. Other perinatal outcomes examined included Apgar scores. Of the 15 included trials, 14 evaluated Apgar scores, and all but one of these found substantially equal scores in the induction and monitoring groups. Dyson, Miller, and Armstrong (1987) reported that a higher proportion of babies in the monitoring group had Apgar scores < 7 at 1 minute (21 percent vs. 11 percent in the induction group); however, similar proportions of infants

in the two groups had scores < 7 at 5 minutes. There is evidence, based on these trials, to conclude that Apgar scores do not change significantly when comparing induction versus monitoring of pregnancies.

Potential maternal benefits. Only one trial (Cardozo, Fysh, and Pearce, 1986) measured patient satisfaction, patient preferences, or quality of life. There were no significant differences in the proportion of patients "pleased" with (49 percent, planned induction; 53 percent, expectant management) or "disappointed" by (15 percent, planned induction; 11 percent, expectant management) their management.

Risks

Perinatal morbidity and mortality. Hyperstimulation of the uterus from induction agents can result in fetal compromise, leading to the need for cesarean section or even fetal death. Because fetal compromise in labor with subsequent need for cesarean section is also associated with prolonged gestation, differences in "risks" for fetal compromise between planned induction and expectant management are the inverse of differences in "benefits" and are discussed above.

Continued fetal growth during expectant management could conceivably lead to an increased risk of macrosomia and shoulder dystocia. In the study by Dyson, Miller, and Armstrong (1987), the proportion of infants with a birthweight greater than 4,000 grams was higher in the expectant management group (28.2 percent) than in the induction group (19.1 percent), though the difference did not reach statistical significance, and no correlation with shoulder dystocia or birth injury was reported. Katz, Yemini, Lancet, et al. (1983) also reported that the incidence of birthweight greater than 4,000 grams was higher in the expectant management group (29.5 percent vs. 7.9 percent; $p < 0.05$), but again no correlation with birth injury was reported. Ohel, Rahav, Rothbart, et al. (1996) found no difference in the proportion of infants with a birthweight greater than 4,000 grams (8.6 percent vs. 8.7 percent). Augensen, Bergsjø, Eikeland, et al. (1987) reported only one case of "difficult shoulder delivery" in the entire study.

In the two large multicenter trials comparing planned induction and expectant management, there were no significant differences in reported rates of macrosomia, shoulder dystocia, or birth injury to the fetus. In the National Institute of Child Health and Human Development (NICHD) Maternal-Fetal Network Trial (National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units, 1994), the incidence of birthweight greater than 4,500 grams was similar in the two induction arms and the expectant management arm, and there was only one case of nerve injury (in one of the induction arms). In the even larger Canadian Multicenter Post-term Pregnancy Trial (Hannah, Hannah, Hellmann, et al., 1992), neither the proportion of infants with a birthweight greater than 4,500 grams (4.6 percent in the induction group vs. 5.5 percent in the expectant management group), nor the incidence of shoulder dystocia (1.4 percent in the induction group vs. 1.6 percent in the expectant group) was significantly different in the two groups.

These results suggest, as would be expected, that continued growth occurs in most infants managed expectantly, resulting in higher proportions of infants over 4,000 grams. Since there is debate as to whether weights between 4,000 and 4,500 grams have any clinical relevance (ACOG, 2000), it is not surprising that there are no reported differences in birth injury. The fact that trials that defined macrosomia as greater than 4,500 grams found no difference in either the proportion of babies weighing more than 4,500 grams or incidence of shoulder dystocia suggests

that elective induction at a predefined gestational age does not have prophylactic benefit—i.e., induction at a given gestational age prior to the development of "macrosomia" does not have an impact on shoulder dystocia.

Cesearean section. Of the 15 included trials, two found a statistically increased risk of overall cesarean section with induction, while three trials found a statistically increased risk of overall cesarean section with expectant monitoring (Table 21).

Meta-analysis and subgroup analyses performed as part of a recent Cochrane review (Crowley, 2000) found no significant differences in cesarean delivery rates in any group or subgroup (Table 22). If anything, cesarean rates tend to be slightly lower in the elective induction groups.

Hannah, et al., published an interesting reanalysis of the Canadian study in 1996 (Hannah, Huh, Hewson, et al., 1996). In this new analysis, women who were randomized to induction or expectant management were stratified based on whether labor was ultimately induced or spontaneous. In the induction arm, 772/1,149 women (67.7 percent) were induced, while 377/1,149 (33.3 percent) went into spontaneous labor prior to scheduled induction. In the expectant management group, 405/1,128 (35.9 percent) were induced for various indications, while 723/1,128 (64.1 percent) went into spontaneous labor. There were no significant differences in cesarean section rates between women randomized to induction who were induced (29.5 percent), women randomized to induction who went into spontaneous labor (25.7 percent), and women who were managed expectantly who went into spontaneous labor (25.7 percent). However, the cesarean section rate was significantly increased in women randomized to expectant management who were induced (42.0 percent). These women were significantly more likely to be nulliparous, to have a closed cervix at the onset of labor, and to have a longer interval from induction to delivery. When compared with the expectantly managed women in spontaneous labor, they had significantly higher cesarean section rates for fetal distress or dystocia; such differences were not seen when the two subgroups in the induction arm were compared.

These differences are consistent with several findings discussed earlier in this report:

- Women whose onset of labor is considerably later than average may represent a distinct subgroup with different physiological characteristics of the uterus and cervix. This is consistent with the higher proportion of women with closed cervices and may also explain the higher rates of cesarean section for dystocia. This also may be related to parity. Presumably, women are included in this group who reach a predefined date for induction without going into spontaneous labor and with normal antepartum testing.

- Provider knowledge of antepartum testing results may affect thresholds for cesarean delivery. It seems likely that providers caring for women whose inductions were indicated because of abnormal antepartum tests would be less tolerant of intrapartum fetal heart rate abnormalities or less likely to tolerate labor progress that was slower than average. This would explain some of the differential rates by indication.

- As Crowley (2000) points out, women induced in the expectant management arm were less likely to receive prostaglandins. This would be a bias in favor of induction. The reanalysis by

Hannah and colleagues (Hannah, Huh, Hewson, et al., 1996) models this based on assumptions about prostaglandin efficacy, and finds that, at worst, there would be no difference in cesarean section rates between groups. In addition, our review of the literature on induction agents (discussed under Question 3) suggests that the effectiveness of prostaglandins in terms of expediting delivery may be proportional to risk of fetal heart rate abnormalities in labor. If this is the case, then any decrease in cesarean section rates for failed induction or dystocia might well be accompanied by an increase in cesarean sections for fetal distress.

In summary, the randomized trial literature consistently shows that elective induction does not result in increased cesarean section rates compared with management strategies based on antepartum testing. If anything, cesarean section rates are slightly lower in women who are electively induced.

Operative vaginal delivery. No studies reported specifically on maternal trauma related to vaginal delivery. Because operative vaginal delivery is clearly associated with an increased risk of maternal injury (Johanson and Menon, 2001), evidence of a difference in the rates of operative vaginal delivery in one group or the other would be suggestive of an increased risk of trauma to the pelvic floor, vagina, or perineum. In seven of the eight studies where this outcome was reported (Bergsjø, Huang, Yu, et al., 1989; Cardozo, Fysh, and Pearce, 1986; Egarter, Kofler, Fitz, et al., 1989; El-Torkey and Grant, 1992; Hannah, Hannah, Hellmann, et al., 1992; Herabutya, Prasertsawat, Tongyai, et al., 1992; Martin, Sessums, Howard, et al., 1989), there were no significant differences between the induction and expectant management groups. In the remaining trial (Hedén, Ingemarsson, Ahlström, et al., 1991), there was a significant difference, with 2.8 percent of the induction group and 15.5 percent of the expectant management group undergoing operative vaginal delivery ($p < 0.01$); the majority of these deliveries in both groups were for "secondary arrest." There are no obvious reasons why the results of this study varied so dramatically from the others. Mean birthweight in the two groups was similar. The standard deviation of the preintervention Bishop score was slightly wider in the expectant management group, and the method of randomization was based on a registration number rather than on randomly generated numbers. One possible explanation for the study's finding on operative vaginal delivery is that the pseudorandomization scheme resulted in some systematic differences in the groups. Another possibility is that use of oxytocin for labor augmentation may have been less aggressive in the expectant management group for some reason.

Overall, the studies reviewed suggest that there is no difference in operative vaginal delivery rates between expectant management and planned induction protocols.

Other maternal risks. There were no differences in the risk of maternal infection or other morbidity in three of the four trials that reported these outcomes (El-Torkey and Grant, 1992; National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units, 1994; Witter and Weitz, 1987). In the remaining, very small trial (Martin, Sessums, Howard, et al., 1989), the proportion of women with "maternal morbidity" was higher in the induction arm (4/12, or 33 percent) than in the expectant management arm (2/10, or 20 percent). No significance testing was reported.

Costs and Resource Use

Direct measures of cost. Only two studies reported direct measures of cost, the Canadian Multicenter Post-term Pregnancy Trial (Hannah, Hannah, Hellmann, et al., 1992) and a smaller study by Witter and Weitz (1987). The Canadian study found that induction of labor was associated with a lower cost compared with monitoring. The mean cost per patient (in 1991 Canadian dollars) of a prolonged pregnancy managed through monitoring was $3,132 (95 percent CI, $3,090 to $3,174), compared with induction, which cost $2,939 (95 percent CI, $2,898 to $2,981) per patient. The difference between the two groups ($193 per patient) was statistically significant. The authors of the study estimated that switching to planned induction could save up to $8 million per year in Canada.

Witter and Weitz (1987) found, on the contrary, that mean costs were higher for planned induction than for monitoring by approximately $250 per patient. This study had a much smaller patient population (n = 200). Because costs frequently are not normally distributed, the effects of a few patients with complications or very long stays may be magnified compared with a larger study.

Indirect measures of resource use. Several studies that did not report direct costs did report outcomes that are indirect measures of resource use, such as overall length of maternal or infant stay in the hospital. The extent to which these results are generalizable is limited, since length of stay varies internationally and has changed dramatically in the United States over recent years. Moreover, overall length of stay may not be entirely related to overall resource use (Tai-Seale, Rodwin, and Wedig, 1999). For women delivering in a hospital, the majority of resource use occurs during the time from admission to delivery, with a sharp decrease after delivery and even further decreases after the first 24 hours. Thus, even if the mean length of stay is equivalent between two groups, the resource use may vary widely depending on what proportion of the time was spent in the delivery suite. In addition, studies that report only hospital use and not outpatient use of resources (for antepartum testing, other office visits, etc.) will not reflect the overall medical costs of a particular strategy. Finally, none of the included studies addressed the nonmedical costs—such as transportation, time lost from work, child care for women with other children, and so on—associated with various strategies for managing prolonged pregnancy.

Table 23 shows reported mean maternal lengths of stay for the six trials where this was reported. There are no obvious trends. Because reporting of the proportion of time spent in labor versus postpartum was minimal, no additional inferences about relative resource use can be drawn.

Only one study (Dyson, Miller, and Armstrong, 1987) reported data on mean neonatal length of stay, with no significant differences between the induction and expectant management groups (3.0 days vs. 3.3 days, respectively).

Tables 24, 25, and 26 summarize perinatal and maternal outcomes and resource use for all trials reviewed.

Methodological Issues

Study Design

All of the included trials were described as "randomized." Four were in fact only pseudorandomized (i.e, treatment was allocated based alternate medical record numbers or birth dates, rather than by randomly generated numbers), which introduces the possibility of bias (Cardozo, Fysh, and Pearce, 1986; Hedén, Ingemarsson, Ahlström, et al., 1991; Katz, Yemini, Lancet, et al., 1983; Ohel, Rahav, Rothbart, et al., 1996). Two studies did not describe the method of randomization used (Egarter, Kofler, Fitz, et al., 1989; Herabutya, Prasertsawat, Tongyai, et al., 1992).

As discussed above and pointed out by Crowley (2000), the practical and ethical difficulties of blinding clinicians to either the target intervention or the results of antepartum testing results in an inherent bias against expectant management. Abnormal antenatal monitoring could influence a clinician's thresholds for performing a cesarean section, either by making the diagnosis of "fetal distress" more likely or by a decreased willingness to augment labor aggressively.

In any trial of planned induction versus expectant management with antepartum testing, a certain proportion of women randomized to planned induction will go into spontaneous labor, while a proportion of women randomized to expectant management will have abnormal antepartum testing results; or, as observed in the Canadian Multicenter Post-term Pregnancy Trial (Hannah, Hannah, Hellmann, et al., 1992), patients or providers may request induction. These subjects are quite correctly analyzed in the groups to which they are randomized, rather than in accordance with the "treatment" received, since the trial is not comparing spontaneous delivery to induction, but instead, management strategies undertaken with the knowledge that some women will deliver spontaneously prior to scheduled induction, and some women will require (or request) induction during expectant management.

Outcome Measurement

All studies reported results for "hard" outcomes such as perinatal mortality and cesarean section rates. Reporting of other outcomes of interest was more variable. Many outcomes are subject to inherent difficulties with reproducibility and bias (e.g., the diagnosis of "fetal distress"), variability in operator preferences and skills (e.g., operative vaginal delivery rates), or are of uncertain long-term clinical significance (e.g., meconium-stained amniotic fluid in the absence of meconium aspiration, or Apgar scores). Other measures, such as patient preferences for different management strategies, longer-term neonatal outcomes, and vaginal and perineal trauma, would be of significant interest to patients, clinicians, and policymakers. We identified one cohort study published in 1991 which showed that patients' preferences for induction versus expectant management changed with advancing gestation: 45 percent of women preferred conservative management at 37 weeks, compared with 31 percent at 41 weeks (Roberts and Young, 1991). Measurement of these preferences in light of data published subsequent to this study, and using methods developed and refined in the past decade, is needed. Detailed measurement of both medical and nonmedical costs is also lacking in the studies reviewed.

Comparability and Generalizability

The gestational age at which interventions were begun, as well as the methods used for induction and monitoring, varied between studies. Because variability in these methods may result in quite different outcomes, caution should be used when comparing outcomes that could possibly be affected by different methods of labor induction (such as cesarean section rates or time spent in labor) or different protocols for fetal monitoring (such as perinatal mortality) between studies. In addition, clinical management decisions may vary between practitioners. Especially in smaller trials, unequal distribution of different practitioners with different preferences and thresholds for management of labor may have resulted in some differences in outcomes.

Readers also must consider the degree to which these studies are generalizable to particular settings. If these methods or protocols are substantially different from those used in a particular setting, then the results may not be applicable. For example, the Canadian Multicenter Post-term Pregnancy Trial did not use prostaglandins for induction of women with abnormal antepartum testing (Crowley, 2000; Hannah, Hannah, Hellmann, et al., 1992). Use of prostaglandins could have changed the results by yielding lower cesarean rates in the induction arm through more successful inductions, as pointed out by Crowley (2000). On the other hand, the use of these agents in women with potentially compromised fetuses could have resulted in even higher cesarean section rates because of fetal compromise. A reanalysis of the Canadian trial using published success rates for prostaglandins found that more liberal use of these agents would still lead to a significantly higher cesarean section rate in the expectant management group because the cesarean section rate in the group induced because of abnormal testing would be substantially higher (Hannah, Huh, Hewson, et al., 1996).

Statistical Issues

Only the Canadian trial (Hannah, Hannah, Hellmann, et al., 1992) was sufficiently powered to detect differences in rare perinatal outcomes. Many of the remaining studies were also under-powered to detect differences in dichotomous outcomes.

Inappropriate summary measures and statistical tests were frequently used (e.g., mean parity or Bishop score, with comparison by t-test, when nonparametric statistics would be more appropriate). Variables that are frequently not normally distributed, such as length of stay and costs, also were not uniformly reported using medians, and the effect of a few outliers on comparisons was not evaluated.

Summary

Despite the methodological issues raised above, there is a consistent finding that perinatal mortality rates are lower with planned induction at 41 weeks or later compared with expectant management, a finding confirmed by a formal Cochrane meta-analysis (Crowley, 2000). Based on the observed absolute risk difference, the Cochrane meta-analysis estimated that 500 inductions were necessary to prevent one perinatal death.

It is interesting to consider these findings in light of our review of antepartum tests under Question 1. We found that there was a consistent qualitative pattern for the majority of tests studied, no matter what surrogate outcome for fetal compromise was used: sensitivity was lower

than specificity, while negative predictive value was higher than positive predictive value. This implies that predictive values are driven by the relatively low rates of adverse outcomes associated with fetal compromise in prolonged pregnancy. If the measures used are valid surrogates for fetal compromise leading to stillbirth, then this should hold true for stillbirth as well: the negative predictive value of antepartum tests for stillbirth should be much greater than the positive predictive value. However, as the risk of stillbirth increases with increasing gestational age after 37 weeks, the negative predictive value should decrease, and the number of stillbirths in the setting of normal test results should increase.

Elective induction of labor results in a lower risk of stillbirth only after 41 weeks. One explanation for this, consistent with the findings on antepartum tests, is that the baseline risk of stillbirth is low enough prior to 41 weeks that the negative predictive value of antepartum tests is quite good. After 41 weeks, the increasing stillbirth risk results in poorer negative predictive value, so that one would expect excess stillbirths compared with elective induction.

Other perinatal outcomes did not appear to differ significantly between induction and expectant management groups.

Maternal outcomes did not differ between women managed with antepartum monitoring or with planned induction with the agents used in these studies. Specifically, overall cesarean section rates did not differ, either globally or in the subgroups analyzed by the Cochrane group (Crowley, 2000). If anything, cesarean section rates were lower in the induced groups.

Only one large trial reported costs, and based on 1992 costs and care provided, planned induction at 41 weeks was less expensive than expectant management with antepartum testing. However, because of significant changes in the technologies used and the economics of medicine in the interim, additional research is needed to better understand the cost implications of these two strategies. For example, if elective induction at 41 weeks is deemed to be preferable from a clinical standpoint for most patients, then a thorough analysis of the resources needed to institute such a policy would have to incorporate factors such as staffing on labor and delivery suites and postpartum units, since temporal patterns of patient flow may change.

Elective induction of labor at 41 weeks consistently appears to reduce the risk of stillbirth compared with management with antepartum testing, with no increase in maternal or neonatal risks, including no increase in cesarean section rates. At least 500 inductions would be needed to prevent one stillbirth. The societal tradeoffs in terms of economic resources used are unclear because of a lack of strong data applicable to current practice. Individual patients may have different values for these outcomes or perhaps for the “process” of childbirth—some women may place a very high value on avoiding any medical intervention.

Question 3: What are the benefits, risks, and costs of currently available interventions for induction of labor?

Approach

The evidence reviewed so far in this report suggests:

- The risk of perinatal death increases with advancing gestational age.
- There is no direct evidence that antepartum surveillance in prolonged gestation reduces perinatal morbidity or mortality. When surrogate measures are used as outcomes, the

consistent pattern of test characteristics for tests used in antepartum surveillance is for poor sensitivity but high negative predictive value, suggesting that false negative test results will become more likely as the underlying risk of adverse outcomes increases with advancing gestational age.

- Randomized trials show a reduction in perinatal mortality in women induced at 41 weeks gestation compared with women followed with antepartum testing, a finding consistent with increasing risk with advancing gestational age and with the observed patterns of test characteristics. Cesarean section rates are not increased in the elective induction arms of these studies.

Given that induction at 41 weeks appears to be effective in reducing mortality, data about the safest and most effective method of induction are needed in order to determine the optimal management strategy.

This section considers interventions designed to induce labor, including prostaglandin E_2 (PGE_2, or dinoprostone) gel (Prepidil®), PGE_2 tablets, PGE_2 insert (Cervidil®), misoprostol tablets, misoprostol gel, oxytocin, mifepristone, membrane sweeping, nipple stimulation, and other treatments. These methods are used either as primary methods of induction or as adjunctive methods in oxytocin induction. We limited our review to studies where the induction method was randomly assigned and compared with either placebo or a different induction method, and where at least some of the subjects were induced for an indication related to prolonged pregnancy. In this section, we also consider active interventions performed in the ambulatory setting at or near term that are designed to reduce the proportion of women reaching "postdates" or "postterm."

In addition to the results of our review, we report summary conclusions based on meta-analyses performed for the Royal College of Obstetricians and Gynaecologists' (RCOG) recent guideline on induction of labor (Royal College of Obstetricians and Gynaecologists, 2001) in collaboration with the Cochrane Collaboration.

Results

Castor Oil

We identified one randomized trial of castor oil used at term to promote spontaneous labor. Garry, Figueroa, Guillaume, et al. (2000) randomized women to 60 mg castor oil given orally in apple or orange juice (n = 52) or no treatment (n = 48). Mean gestational age was 284.4 ± 4.2 days in the castor oil group and 284.7 ± 3.6 days in the no treatment group. In the castor oil group, 57.7 percent of the subjects were in labor within 24 hours compared with 4.2 percent in the no treatment group ($p < 0.001$). Cesarean section rates were 19.2 percent in the castor oil group and 8.3 percent in the no treatment group ($p = 0.20$), but the study was underpowered to detect this difference or differences in rare outcomes such as uterine rupture. Of note, all women in the castor oil group experienced nausea. Other outcomes, such as proportion of women induced for other reasons or neonatal outcomes, were not reported.

The RCOG guideline (Royal College of Obstetricians and Gynaecologists, 2001) did not address castor oil. The most recent Cochrane review on the topic (Kelly, Kavanagh, and Thomas,

2001) identified the article cited above (Garry, Figueroa, Guillaume, et al., 2000) and reached conclusions similar to our own.

Breast Stimulation

We identified two studies that evaluated the use of breast stimulation in promoting the onset of labor near term and one that evaluated breast stimulation as a method of induction. Elliot and Flaherty (1984) randomized 100 women to either breast stimulation (manual stimulation of the nipple and areola for 15 minutes, alternating breasts, for a total of 1 hour at a time, three times daily) beginning at 39 weeks or a control pelvic examination; women in the control group were asked to abstain from sexual intercourse and avoid breast stimulation. Both groups were reevaluated at 42 weeks. Women with Bishop scores of 8 or greater were induced; others were followed with contraction stress tests. Five women in the breast stimulation group reached 42 weeks, compared with 17 in the control group; significance testing was not performed. Women in the breast stimulation group were significantly less likely to be induced after 42 weeks. The study was underpowered to detect differences in important outcomes, especially for the subgroup of women beyond 42 weeks.

Kadar, Tapp, and Wong (1990) randomized women at 39 weeks to either daily unilateral manual nipple stimulation "for as long as was practically feasible" (n = 60) or to no nipple stimulation (n = 76). There were no significant differences in any of the outcomes reported, including the proportion going into spontaneous labor, postterm deliveries, or median duration of pregnancy. Survival analysis showed that duration of pregnancy was related only to gestational age at enrollment and Bishop score. The authors also noted that adherence to the prescribed regimen was poor: 70 percent of the women assigned to the nipple stimulation group either failed to perform nipple stimulation at all or did so for less than 2 hours total during the entire study.

Chayen, et al., compared nipple stimulation using an electric breast pump to oxytocin as a method of induction (Chayen, Tejani, and Verma, 1986). In this study, only 29 percent of the inductions were for prolonged pregnancy. Thirty subjects were induced initially with a breast pump, while 32 received oxytocin. Time to achieve regular contractions and adequate labor as documented by intrauterine catheter were significantly less in the breast pump group. Cesarean section rates were also lower (26.7 percent vs. 43.7 percent in the oxytocin group), although this difference was not significant. Patients in the oxytocin group were more likely to have a higher Bishop score at baseline. Results were not reported separately by parity or for the subgroup of women induced for prolonged pregnancy.

In summary, because of lack of significance testing, poor compliance, or lack of power, the available randomized trials do not allow conclusions to be drawn about the effectiveness of breast stimulation in promoting labor or as a method of induction. The RCOG guideline (Royal College of Obstetricians and Gynaecologists, 2001) did not address this topic.

Relaxin

We identified three randomized trials of relaxin. Evans, Dougan, Moawad, et al. (1983) randomized women at 41 weeks gestation scheduled to undergo oxytocin induction of labor to intracervical or vaginal insertion of 4 mg relaxin (n = 10), 2 mg relaxin (n = 13), or placebo (n = 14); if the patient reached 42 weeks gestation, then labor was induced. No significant differences in any parameters, including days to admission, spontaneous labor, or time to

delivery, were noted. There were trends towards a shorter time to delivery in the relaxin groups, but the study was underpowered to detect a difference for this outcome.

Bell, Permezel, MacLennan, et al. (1993) randomized women scheduled for induction for prolonged pregnancy to intravaginal 1.5 mg recombinant human relaxin (n = 18) or placebo (n = 22). No significant differences in any outcomes were reported. The authors noted that a low dose was deliberately chosen to help establish a safety profile for relaxin.

` Brennand, et al., randomized women between 37 and 42 weeks, "most" of whom were being induced for pregnancy-induced hypertension or prolonged pregnancy, to placebo or 1 mg, 2 mg, or 4 mg of recombinant relaxin (Brennand, Calder, Leitch, et al., 1997). There were no significant differences in any outcome except for slightly elevated baseline fetal heart rates after relaxin.

In summary, there are insufficient data available on relaxin to draw any conclusions about its safety or efficacy in induction of labor in women with prolonged pregnancy.

Sweeping of the Membranes

We identified 12 trials evaluating the efficacy of sweeping (or "stripping") of the membranes, 11 designed to evaluate the use of this intervention to promote spontaneous labor and reduce the need for induction and one in which it was used as a method of induction. In general, sweeping the membranes involves inserting a finger into the cervix and rotating the finger in the plane between the fetal membranes and the cervix and lower uterine segment. Details of the techniques used varied between studies and are described for each study in Evidence Table 3. Table 27 summarizes the 11 trials of membrane sweeping as a labor promoter.

All studies except one consistently showed higher rates of labor within a predefined time period, usually 1 week, in women randomized to active membrane sweeping. The proportion of women induced was also consistently lower in groups randomized to membrane sweeping. No differences in adverse outcomes, including infection or bleeding, were noted in any study. Level of patient discomfort during the procedure was not assessed in any study.

The one study that did not show a difference in outcomes (Crane, Bennett, Young, et al., 1997) was different from the other trials in several ways. Membrane stripping was performed only once. Patients in the stripping group were more likely to be nulliparous and to have lower Bishop scores. Stratified analyses and logistic regression did not show significant effects, but it is possible that the smaller sample size in these subgroups limited power. In addition, a survival analysis showed a decrease in the median time from enrollment to delivery (6.5 days for stripping, compared with 8 days for controls), but this difference was not significant.

In the one study in which membrane sweeping was used as an adjunct to induction of labor, Boulvain, et al., randomized women to sweeping of the membranes (n = 99) or vaginal examination only (n = 99) prior to induction of labor for "nonurgent" indications (Boulvain, Fraser, Marcoux, et al., 1998). Eighty-five percent of the patient population was induced for prolonged pregnancy. Mean time from randomization to onset of labor was significantly shorter in the sweeping group (76 hours vs. 98 hours; p = 0.01), but no significant differences were seen in other outcomes except patient discomfort (odds ratio [stripping vs. control], 2.52; 95 percent confidence interval [CI], 1.60 to 3.99), bleeding, and painful contractions without labor.

In summary, in all but one study, sweeping the membranes consistently promoted labor at term and reduced the incidence of induction for prolonged pregnancy. As with the majority of the interventions reviewed in this report, there are no data on patient preferences for this

intervention. One study found that women who undergo membrane stripping are more likely to experience discomfort, bleeding, and painful contractions without labor compared with controls. Another issue is that the majority of studies excluded women whose cervices would not allow introduction of the examiner's finger; thus, the conclusions described are applicable only to those pregnant women at term whose cervices are dilated enough to allow introduction of an examiner's finger.

Similar findings have been reported in a Cochrane review (Boulvain and Irion, 2001) and incorporated into the RCOG guidelines (Royal College of Obstetricians and Gynaecologists, 2001).

Mechanical Devices

We identified two randomized trials of the use of mechanical devices such as Foley catheters, which are inserted into the cervix and then inflated. Atad, et al. (Atad, Hallak, Auslender, et al., 1996) compared 3 mg PGE_2 gel (n = 30), oxytocin (n = 30), and a double-balloon catheter invented by one of the investigators (n = 35). Patients in the first two groups crossed over to the catheter arm if the Bishop score was ≤ 4 at 12 hours, while patients in the catheter group received PGE_2 if the Bishop score was ≤ 4 at 12 hours. More patients in the catheter group had cervical dilation > 3 cm after 12 hours (86 percent vs. 23 percent in the oxytocin group and 50 percent in the PGE_2 group; $p < 0.01$). Both PGE_2 and the balloon device had higher rates of vaginal delivery (PGE_2, 70 percent; catheter, 77 percent; oxytocin, 27 percent) and lower rates of cesarean section among patients with cervical dilation after the initial intervention (PGE_2, 13 percent; catheter, 18 percent; oxytocin, 43 percent). Only 18 percent of the inductions in this study were for prolonged pregnancy.

Sciscione, et al., randomized 53 women to misoprostol and 58 to mechanical dilation with a 16 F Foley catheter with a 30 cc balloon (Sciscione, Nguyen, Manley, et al., 2001). There were no significant differences in change in Bishop score, vaginal delivery rates, or time to delivery in the two groups. Uterine tachysystole and passage of meconium were significantly more frequent in the misoprostol group. There was a trend towards higher cesarean section rates for nonreassuring fetal heart rate tracing in the misoprostol group (24 percent vs. 12 percent; $p = 0.09$), in a study where the sample size was determined based on change in Bishop score. Only 16 of 111 women in this study were induced for an indication of prolonged pregnancy.

In these two trials, mechanical devices appear to be comparable to prostaglandins in terms of delivery success, with lower rates of fetal heart rate tracing changes associated with frequent uterine contractions. As with membrane sweeping, applicability is limited to women whose cervix is dilated enough to allow introduction of a catheter. As with the majority of the other interventions reviewed, these studies also included relatively few women in the population of interest (prolonged pregnancy with no other risk factors) and were underpowered to detect differences in many important outcomes.

Mechanical devices alone are not addressed specifically in published Cochrane reviews or in the RCOG guideline (Royal College of Obstetricians and Gynaecologists, 2001).

Oxytocin Dosing

We identified one randomized trial comparing two dosing regimens of oxytocin. Satin, Hankins, and Yeomans (1991) randomized women being induced for prolonged pregnancy to a

"slow-dose" regimen (an initial dose of 2 mU/min, with increments of 1 mU/min at 30-minute intervals) or a "fast-dose" regimen (an initial dose of 2 mU minute with increases of 2 mU/min at 15-minute intervals). Induction failure was more likely in the slow-dose group (31 percent vs. 8 percent; $p < 0.05$). Time to delivery was shorter in the fast-dose group in both nulliparous women (9 hours vs. 15 hours; $p < 0.05$) and multiparous women (8 hours vs. 11 hours; $p < 0.05$). No significant differences were observed in other outcomes. There was a trend towards more hyperstimulation episodes requiring cessation of oxytocin in the fast-dose group, but the study was underpowered to detect a difference.

There is no formal comparison of oxytocin dosing regimens in published Cochrane reviews. The RCOG guideline development group reviewed dosing regimens in 11 trials of oxytocin with and without amniotomy. Their qualitative conclusions were: (1) lower dose regimens were not associated with an increase in operative delivery rates; (2) regimens with incremental rises in dose more frequently than every 30 minutes were associated with an increase in uterine hypercontractility; (3) lower dose regimens were not associated with an increase in specified delivery intervals; and 4) higher dose regimens were associated with an increase in the incidence of precipitous labor (Royal College of Obstetricians and Gynaecologists, 2001).

Prostaglandins

Of the randomized trials identified, 20 evaluated PGE_2 (dinoprostone) gel, five evaluated PGE_2 tablets, one evaluated the Cervidil® insert, one evaluated low-dose (2 mg) PGE_2 vaginal suppositories, and 22 examined misoprostol. Placement of the prostaglandin was either intravaginal (usually in the posterior fornix) or intracervical. The site of application is described for each study in Evidence Table 3 and in the text below.

PGE_2 gel in an ambulatory setting to reduce the need for induction. Five studies examined the effect of PGE_2 gel versus placebo (Buttino and Garite, 1990; Doany and McCarty, 1997; Lien, Morgan, Garite, et al., 1998; O'Brien, Mercer, Cleary, et al., 1995; Sawai, Williams, O'Brien, et al., 1991). Doany and McCarty (1997) randomized patients to one of four arms: (1) no membrane stripping and placebo gel; (2) no membrane stripping and PGE_2 gel; (3) membrane stripping and placebo gel; or (4) membrane stripping and PGE_2 gel. Gel was placed in the posterior vaginal fornix. PGE_2 gel without membrane stripping was not significantly different from placebo without stripping for any outcome. All patients in this study were 41 weeks or greater in gestational age.

Lien, et al., a randomized trial of intracervical PGE_2 gel ($n = 43$) versus placebo ($n = 47$) begun after 40 weeks, found no significant differences between the two arms in the interval from admission to delivery, cesarean sections, or maximum oxytocin dosage (Lien, Morgan, Garite, et al., 1998). For patients who presented with a Bishop score between 3 and 6, those who were randomized to PGE_2 gel were less likely to be induced than those treated with placebo gel.

Sawai, Williams, O'Brien, et al. (1991) randomized women at 41 weeks to either weekly PGE_2 gel in the posterior fornix ($n = 24$) or weekly placebo gel. Induction occurred if the Bishop score was greater than 9, in the event of abnormal fetal heart rate testing, or at 44 weeks. There were no significant differences in neonatal outcomes, cesarean section rates, length of labor, or time from randomization to admission between the two groups, but the study was underpowered to identify differences in most categorical variables.

Buttino and Garite (1990) randomized women at 41-6/7 weeks to either intracervical PGE_2 (n = 23) or placebo (n = 20). There were no significant differences in any outcome, including neonatal outcomes, cesarean section rate, or time to delivery. Cesarean section rates were lower in the PGE_2 group (21.7 percent vs. 35.0 percent), but the study was underpowered to detect a difference. Gestational age at delivery and time from randomization to delivery were not significantly different in the two induction groups.

O'Brien, et al., randomized women at 38-39 weeks to intravaginal PGE_2 gel (n = 50) or placebo (n = 50) daily for 5 days (O'Brien, Mercer, Cleary, et al., 1995). PGE_2 gel resulted in significantly fewer pregnancies going beyond 40 weeks (40 percent vs. 66 percent; $p < 0.016$), although not in the proportion of pregnancies reaching 42 weeks (4 percent vs. 6 percent). Induction rates were lower in the PGE_2 group (12 percent vs. 28 percent; $p = 0.08$).

PGE_2 gel as an adjunct to oxytocin. A randomized trial conducted by the National Institute of Child Health and Human Development (NICHD) Network of Maternal-Fetal Medicine Units (1994) compared induction between 41 and 42 weeks and expectant management. The induction group in this trial was split into two arms: intracervical PGE_2 gel plus oxytocin (n = 174) and placebo gel plus oxytocin (n = 174). No significant differences in neonatal or maternal outcomes, including cesarean section rates, were detected between the two groups. Sample size estimates for this trial were based on perinatal morbidity and mortality and maternal mortality.

Rayburn, et al., compared intracervical PGE_2 gel (n = 55) to placebo (n = 63) prior to induction of labor with oxytocin at 42 weeks (Rayburn, Gosen, Ramadei, et al., 1988). Overall cesarean section rates (18 percent with PGE_2 gel vs. 33 percent with placebo; $p < 0.05$) and mean time to delivery (5.5 hours vs. 9.5 hours with placebo; $p < 0.01$) were significantly lower with PGE_2 gel.

Chatterjee, et al., compared 2 mg PGE_2 gel to placebo (Chatterjee, Ramchandran, Ferlita, et al., 1991). Bishop scores were significantly improved in patients receiving the active gel; the study was underpowered to detect any other differences.

PGE_2 gel dosing. Voss, Cumminsky, Cook et al. (1996) compared the use of intracervical PGE_2 gel in three different dosing regimens: 0.125 mg (n = 79), 0.25 mg (n = 70), and 0.5 mg (n = 80). For each of the outcomes described (fetal heart rate abnormality, cesarean sections, mean change in Bishop score, hyperstimulation, and time to active phase labor/complete dilation/delivery), there was no significant difference noted for the various doses of PGE_2 gel. Only 31 percent of subjects in this study were induced for prolonged pregnancy.

MacKenzie and Burns (1997) compared a single vaginal dose of 2 mg PGE_2 gel, with amniotomy and oxytocin if no labor occurred within 14-20 hours of treatment, with 2 mg of PGE_2, followed by a second application in 6 hours if no labor occurred or if the Bishop score was less than 9. Sixty-eight percent of the patients in this trial were induced for prolonged pregnancy. The only significant difference noted was a shorter time to delivery in the two-dose group among multiparous women (mean 785 minutes vs. 927 minutes in the single-dose group).

Graves, et al., compared PGE_2 gel in doses of 1 mg, 2 mg, and 3 mg to placebo prior to induction with oxytocin (Graves, Baskett, Gray, et al., 1985). Eighteen percent of the inductions were for prolonged pregnancy. There was a significant increase in Bishop score after the active gel compared with placebo, but this effect was not dose-related. There was a dose-related increase in the proportion of women entering spontaneous labor after insertion of the gel. There was a trend toward more uterine hypercontractility with higher doses of the gel, although the

study was underpowered to detect a significant difference. Other outcomes were not significantly different between the active and placebo groups, although the study lacked power to detect many differences.

PGE_2 gel versus PGE_2 tablets. One study compared 3 mg PGE_2 tablets to 2 mg PGE_2 gel (Mahmood, 1989). The gel formulation required fewer applications and resulted in greater changes in Bishop score and shorter time to onset of labor than did tablets.

PGE_2 gel versus oxytocin. Two studies were identified that compared the administration of PGE_2 gel to induction by oxytocin infusion. In the first study (Papageorgiou, Tsionou, Minaretzis, et al., 1992), cesarean section for cephalopelvic disproportion and fetal distress, vacuum suction, and hyperstimulation were not statistically different in women randomized to intracervical PGE_2 (n = 83) or oxytocin (n = 82) for induction of labor after 41 weeks. Two outcomes did show benefit to the use of PGE_2 gel. First, babies were less likely to have an Apgar score < 7 at 5 minutes when the cervices of the mother were ripened by PGE_2 gel as opposed to those induced with oxytocin. Also, patients were more likely to be delivered vaginally if ripened by PGE_2 gel (89 percent vs. 71 percent). All subjects in this study had a gestational age of at least 41 weeks.

The second study (Misra and Vavre, 1994) compared administration of intracervical PGE_2 gel (n = 80) with oxytocin (n = 72). Rates of cesarean deliveries were decreased with PGE_2 in primigravidas only (26.3 percent with PGE_2 vs. 47.2 percent with oxytocin; $p < 0.01$). Women in this study were induced for a variety of indications, with a mean gestational age less than 40 weeks.

Placement of PGE_2 gel. One study examined the effect of placement of PGE_2 gel in the posterior vaginal fornix versus in the endocervical canal (Kemp, Winkler, and Rath, 2000). The outcomes that showed significance indicated that patients who received gel administered in the posterior vaginal fornix were more likely to deliver earlier (15.7 hours vs. 19.1 hours) and more likely to deliver in 24 hours (81.6 percent vs. 67.8 percent). In this study, 32.9 percent of the posterior fornix group were induced for prolonged pregnancy (more than 10 days past the estimated date of confinement), and 29.2 percent of the intracervical group were 10 days beyond term.

PGE_2 gel versus membrane stripping. Two studies compared outcomes between PGE_2 gel administration and membrane stripping. In Magann, et al., three groups were randomly assigned to treatment at 41 weeks (Magann, Chauhan, Nevils, et al., 1998). One group received daily intracervical administration of PGE_2 gel, another received daily membrane stripping, and the third group received a daily "gentle cervical examination." Patients in all three groups were induced if the Bishop score became ≥ 8, or at 42 weeks. Inductions at 42 weeks were significantly lower in the two active treatment groups (17 percent in the sweeping group and 20 percent in the PGE_2 group, compared with 60 percent in the controls). Cesarean section rates were higher in the PGE_2 group (8/35, or 23 percent, vs. 5/35, or 14 percent, in the other two groups), a relative risk of 1.6 (95 percent CI, 0.58 to 4.41).

In Doany and McCarty (1997), the effects of membrane stripping, PGE_2 gel (placed in the posterior vaginal fornix), and a combination of the two therapies were evaluated. Patients were randomized at 41 weeks to one of 4 groups: (1) membrane stripping and placebo gel;

(2) membrane stripping and PGE_2 gel; (3) "control" cervical exams and placebo gel; or (4) "control" exams and PGE_2 gel. Gestational age at delivery was significantly lower in the group with both active treatments (median, 290 days vs. 294 days in the two groups with one placebo and 297 days in the group with two placebos; $p = 0.005$). There was a trend towards a higher cesarean rate in the group with both active treatments (11 percent versus 8 percent in the two single-agent arms and 4 percent in the double-placebo group; $p = 0.08$).

These two studies suggest that PGE_2 is equivalent to membrane stripping in terms of promoting labor. In both studies, PGE_2 was associated with higher cesarean section rates, although these differences were not statistically significant. Larger studies would be needed to detect a difference in cesarean rates.

PGE_2 inserts. Only one study was identified that examined the efficacy of the Cervidil® vaginal insert (Wing, Ortiz-Omphroy, and Paul, 1997). This trial compared the Cervidil® insert (10 mg in a timed-release preparation) to 25 µg of misoprostol administered every 4 hours to a maximum of six doses. There were no significant differences between the two groups in neonatal or maternal outcomes. While the mean time to delivery was the same between the two groups, the misoprostol dosing every 4 hours showed a lower rate of tachysystole than the Cervidil® insert.

PGE_2 suppositories. One study evaluated the use of 2 mg intravaginal PGE_2 suppositories ($n = 38$) versus placebo suppositories ($n = 42$) self-administered by the patient on an outpatient basis beginning at 41 weeks (Sawai, O'Brien, Mastrogiannis, et al., 1994). The patients in the PGE_2 arm used fewer suppositories and were admitted for delivery at earlier gestational ages. This resulted in lower antepartum testing charges (mean $477 vs. $647 with placebo; $p = 0.001$). There was a trend towards lower cesarean section rates in the PGE_2 group (2.6 percent vs. 14.3 percent in the placebo group), although this difference was not significant.

In summary, vaginal or intracervical PGE_2 was consistently more effective in achieving cervical ripening or delivery within a specified time period compared with placebo or oxytocin. Cesarean section rates were lower or similar in women treated with PGE_2. There were no differences in perinatal or maternal morbidity or mortality.

Similar findings were reported in the review conducted for the RCOG guideline group. Based on their "conflated" analysis of trials comparing PGE_2 with oxytocin with or without amniotomy, the guidelines recommended PGE_2 as the treatment of choice for induction in women with intact membranes (Royal College of Obstetricians and Gynaecologists, 2001).

Misoprostol

Misoprostol tablets versus placebo. Only one study was identified that compared misoprostol with placebo prior to scheduled induction (Fletcher, Mitchell, Simeon, et al., 1993). A dose of 100 µg misoprostol ($n = 32$) was found to be more effective than placebo ($n = 31$). Time from induction to delivery was lower with misoprostol (22 hours vs. 32 hours), as was cesarean section rate (3 percent vs. 10 percent), although these differences were not statistically significant. The mean Bishop score was increased for patients treated with misoprostol. Only one-third of the randomized patients were induced for prolonged pregnancy.

Misoprostol tablets versus PGE_2 gel. Table 28 summarizes results from the 10 studies that compared intravaginal misoprostol tablets with intracervical or intravaginal PGE_2 gel (Buser, Mora, and Arias, 1997; Chuck and Huffaker, 1995; Fletcher, Mitchell, Frederick, et al., 1994; Gottschall, Borgida, Mihalek, et al., 1997; Herabutya, Prasertsawat, and Pokpirom, 1997; Howarth, Funk, Steytler, et al., 1996; Kadanali, Küçüközkan, Zor, et al., 1996; Mundle and Young, 1996; Varaklis, Gumina, and Stubblefield, 1995; Wing, Jones, Rahall, et al., 1995).

The studies examined a range of doses and frequency of dosing with similar results. The time from induction to delivery was consistently shorter in patients treated with misoprostol, both for all patients and for those with vaginal delivery. With one exception, misoprostol was shown to cause higher frequency of uterine hyperstimulation, hypertonus, or tachysystole, although studies were often underpowered to detect significant differences in these outcomes. All studies indicated that misoprostol was an effective agent for cervical ripening and induction, often more effective than PGE_2 gel, and showed no significant difference in the rates of cesarean section. One study (Buser, Mora, and Arias, 1997) showed an increase in cesarean section rates for patients treated with misoprostol; this was attributable to significantly higher rates of nonreassuring fetal heart rate patterns. Of note, the majority of subjects in these studies were not women being induced for prolonged pregnancy.

Misoprostol dosing studies. Two studies evaluated various dosing regimens for misoprostol. In Farah, et al., intravaginal administration of doses of 25 µg versus 50 µg every 3 hours was evaluated (Farah, Sanchez-Ramos, Rosa, et al., 1997). In this study, the incidences of hyperstimulation, tachysystole, and cord pH < 7.16 were greater in patients on the 50-µg regimen. In comparison, patients given 50 µg every 3 hours were more likely to have shorter start-to-delivery times and more vaginal deliveries.

In Wing and Paul (1996), the dosing regimen was 25 µg given either every 3 or 6 hours. Patients randomized to the 6-hour regimen had longer times to delivery, more frequently required oxytocin augmentation, and had more failed inductions than those on the 3-hour regimen.

Misoprostol versus oxytocin. Three studies compared the effect of intravenous oxytocin with intravaginal misoprostol (Escudero and Contreras, 1997; Kramer, Gilson, Morrison, et al., 1997; Sanchez-Ramos, Kaunitz, Del Valle, et al., 1993). Although the studies used varying dosages of misoprostol, the conclusions were similar. Patients treated with misoprostol had shorter induction-to-delivery times, more vaginal deliveries, and fewer cesarean deliveries for dystocia. Most studies also indicated that higher rates of uterine tachysystole were associated with misoprostol, and studies with higher doses of misoprostol had higher rates of tachysystole. Kramer, et al., found that patients treated with misoprostol also were less likely to use epidural anesthesia, and the costs associated with misoprostol induction were less than for patients induced by oxytocin (Kramer, Gilson, Morrison, et al., 1997). In this study, the costs associated with misoprostol treatment often excluded the cost of epidural anesthesia, longer length of stay (associated with induction), and fewer cesarean deliveries.

Method of delivery with misoprostol. Two studies examined the effect of various methods of delivery for the dosing of misoprostol. Srisomboon, et al., evaluated the effect of 100 µg of misoprostol given intracervically versus intravaginally (after dissolution of the misoprostol pill into an inert gel) (Srisomboon, Piyamongkol, and Aiewsakul, 1997). There were no significant

differences found between the two methods of administration in terms of change in Bishop score, interval from administration to delivery, route of delivery, or perinatal outcome. Rates of uterine tachysystole were similar in the two groups. This study noted that spillage of gel out of the cervix was observed in 70 percent of patients receiving intracervical misoprostol. The investigators concluded that the rates of efficacy between the two methods were similar, and that intravaginal administration was more convenient. Thirty-four percent of the inductions in this study were for prolonged gestation.

Toppozada, Anwar, Hassan, et al. (1997) evaluated the effects of oral versus vaginal misoprostol. Forty patients were randomized to 100 µg every 3 hours administered via the oral or vaginal route. Patients were more likely to be induced successfully via the vaginal route in a shorter interval at a lower dose but were also more likely to experience abnormal fetal heart rate patterns and higher rates of uterine hyperstimulation. The proportion of subjects induced for prolonged pregnancy was not reported in this study.

Misoprostol tablet versus PGE_2 tablet. Four studies were identified that evaluated the effects of intravaginal PGE_2 tablets to intravaginal misoprostol tablets (Chang and Chang, 1997; Fletcher, Mitchell, Frederick, et al., 1994; Lee, 1997; Surbek, Boesiger, Hoesli, et al., 1997). While the dosing regimens for the studies differed, the conclusions were similar. Patients treated with misoprostol were found to have shorter intervals between insertion and delivery, had higher mean Bishop scores 12 hours after administration, and were more likely to deliver in 24 hours. Three of the four studies concluded that misoprostol was a more effective and efficient drug for induction than PGE_2. No significant differences in perinatal outcomes were noted.

Misoprostol versus PGE_2 insert (Cervidil®). One study compared the effects of the Cervidil® vaginal insert with misoprostol (Wing, Ortiz-Omphroy, and Paul, 1997). Patients randomized to treatment with Cervidil® had higher rates of tachysystole and abnormal fetal heart rate patterns. There were no significant differences in perinatal outcomes. Patients treated with misoprostol had shorter intervals from start to delivery than those treated with Cervidil®, but this difference was not significant. This study concluded that misoprostol was as effective as Cervidil®, but that the incidence of uterine tachysystole was significantly lower with misoprostol.

In summary, the majority of the randomized trials of misoprostol showed that misoprostol was more effective in achieving vaginal delivery within 24 hours than were other induction agents. However, misoprostol was also more likely to result in uterine hypercontractility, a not unsurprising correlate of efficacy. All the studies reviewed were underpowered to detect clinically relevant differences in many important outcomes, particularly those having to do with safety. Similar conclusions have been reached by recent Cochrane reviews on misoprostol (Alfirevic, Howarth, and Gaussmann, 2000; Hofmeyr and Gulmezoglu, 2001).

Mifepristone

We identified five studies that compared the efficacy of the progesterone receptor antagonist mifepristone (RU-486) to placebo. Unlike many of the studies discussed above, three of the five focused on patients primarily induced for prolonged pregnancy. All five studies indicated that mifepristone was effective in ripening the cervix. Wing, et al., using 200 mg mifepristone, found significantly more deliveries and vaginal deliveries within 48 hours and a shorter time to delivery with mifepristone compared with placebo; subgroup analysis showed that these effects were

primarily due to the effect in nulliparas (Wing, Fassett, and Mishell, 2000). There were trends towards more abnormal fetal heart rate tracings in labor and more infants with Apgar scores less than 7 at 1 and 5 minutes in the mifepristone group, but these trends did not reach statistical significance.

Three studies evaluated patients who were treated with 400 mg mifepristone versus placebo. In Stenlund, Ekman, Aedo, et al. (1999), the time to onset of labor was shorter and the proportion of patients in labor within 48 hours was significantly greater (81.8 percent vs. 27.3 percent) in the mifepristone group. Median Apgar scores at 1 minute were lower in the mifepristone group, but there were no differences in Apgar scores at 5 or 10 minutes. With only 36 subjects, this study was underpowered to detect differences in many outcomes.

In Giacalone, et al., time to onset of labor and time to vaginal delivery were significantly shorter in the mifepristone group (Giacalone, Targosz, Laffargue, et al., 1998). There were trends towards lower Apgar scores at 1 minute and lower cord pH values, but these were nonsignificant; again, the study was severely underpowered to detect differences in many important clinical outcomes, including cesarean section rate.

In Frydman, et al., the proportion of women going into spontaneous labor, the proportion with Bishop scores less than 4 at presentation for induction, and the mean randomization-to-delivery time were all significantly less in the mifepristone group (Frydman, Lelaidier, Baton-Saint-Mleux, et al., 1992). There were no significant differences in other outcomes and no other trends. Again, the study was underpowered to detect differences in safety-related outcomes. Forty-eight percent of the patients were induced for “postdate” pregnancy.

Elliott, et al., performed a dose-response study comparing placebo with 50 mg and 200 mg of mifepristone in nulliparous women, the “majority” of whom were being induced for prolonged pregnancy (Elliott, Brennand, and Calder, 1998). When a combined outcome measure of either spontaneous labor within 4 days or Bishop score of ≥ 6 at induction was used as the measure of efficacy, there were significant improvements with mifepristone in a dose-related manner. However, mifepristone was also associated in a dose-related manner with significantly more cases of fetal distress in labor and neonatal jaundice. In addition, cesarean rates were significantly lower with 50 mg of mifepristone than with placebo but higher with 200 mg than with placebo ($p = 0.07$), a difference that appears to be attributable to a higher incidence of cesarean delivery for fetal distress in the 200-mg group.

In summary, mifepristone appears to be superior to placebo in terms of achieving labor or cervical ripening within a specified time, but there are consistent trends towards fetal compromise during labor in women who receive mifepristone. Inadequate power to detect potentially important differences in safety argue against the use of mifepristone for induction of labor in prolonged pregnancy outside of research protocols at the present time.

A Cochrane review on this topic found similar evidence of efficacy (Neilson, 2001). Neonatal outcomes were not reported in enough studies to allow conclusions about safety.

Methodological Issues

In reviewing the literature on induction agents, numerous methodological problems consistently reduced our ability to draw conclusions about the benefits and risks of these agents in managing women with prolonged pregnancy. Some of these problems concerned study design; others related to statistical issues.

The following observations may be made about study design:

- Patient population: The majority of the studies evaluating the efficacy of different interventions for induction of labor included subjects with a range of indications for induction and did not report results separately for those women induced because of prolonged pregnancy. This has several implications. First, it is possible that the responsiveness of the uterus and cervix (even with comparable Bishop scores) to a given agent might be quite different between a woman at 37 weeks with preeclampsia and a woman at 42 weeks with no medical complications, leading to different estimates of efficacy. Second, risks for fetal compromise might also be quite different between a woman at 37 weeks with preeclampsia compared with a woman at 41 weeks with no medical complications compared with a woman at 42 weeks with oligohydramnios. The two groups of interest in this report are women induced solely because of prolonged gestation and women induced because of abnormal antepartum surveillance in prolonged gestation. The majority of the literature does not allow us to draw conclusions about the risks and benefits of particular induction agents in these two groups. Several studies also noted differences in outcomes between nulliparous and parous women; the majority failed to stratify results by parity.

- Choice of primary outcomes: Of those studies that stated an a priori sample size estimation, most based it on time-related outcomes, such as time to delivery, time to vaginal delivery, or proportion of subjects delivering within 24 or 48 hours. Although these certainly are important outcomes, sample size estimates based on these types of outcomes will inevitably lead to studies that are underpowered to detect clinically relevant differences in other important outcomes, such as perinatal morbidity or cesarean section rates. This was found throughout the misoprostol literature, where there were consistent trends towards higher rates of uterine tachysystole, hyperstimulation, and nonreassuring fetal heart rate tracings, but most studies were underpowered to detect the differences. Studies that based their sample size estimates on changes in the Bishop score failed to account for the inherent intra- and interobserver variability of this measurement; accounting for this would have led to larger sample sizes.

- Variability in clinical management: As with most of the studies reviewed for this report, variability in clinical management of labor may have resulted in differences in many outcomes, especially cesarean section rates, which make comparisons across studies difficult.

- Patient preferences: Consistently, time to delivery was chosen as an important outcome variable. Not surprisingly, more rapid times to delivery were associated with intermediate markers of fetal compromise or potential fetal compromise. Time to delivery is an important resource use issue. However, given the potential tradeoffs, collection of patient-oriented outcomes (preferences for the tradeoff of time in labor vs. risk of fetal compromise, for example) would be a valuable adjunct to these studies.

- Cost data: Few studies reported cost data. Those that did frequently failed to account for all medical costs and focused only on pharmacy-related costs. This lack of data prevents estimation of cost-effectiveness.

The following observations are made about statistical issues:

- Sample size: As stated above, the choice of primary outcome variable often inhibited the ability of trials to detect potentially clinically relevant differences in important outcomes. This is particularly true for rare but clinically important outcomes such as uterine rupture. There are case reports of uterine rupture occurring in women without previous uterine surgery after induction with misoprostol (Bennett, 1997; Blanchette, Nayak, and Erasmus, 1999); whether the risk of this event is higher in women induced with misoprostol compared with other medications is unclear, since denominator data are not available. However, the lack of statistical power to detect categorical events in the majority of randomized trials of induction agents is a major limitation to interpretation of this literature.

- Choice of statistical tests: Inappropriate statistical tests (e.g., means for integer variables such as parity, Apgar or Bishop score, or for nonnormally distributed variables, such as length of stay or time in labor) were frequently used. Use of these summary measures could potentially lead to false conclusions about the comparability of groups at either baseline or after intervention.

Summary

Based on the above review, we conclude the following:

- The majority of randomized trials of induction agents where a priori sample size estimates were performed are powered based on detecting a difference in outcomes such as time to delivery. This results in a lack of power to detect clinically meaningful differences in categorical outcomes that are less common. This lack of power precludes drawing definite conclusions about the relative safety of different agents.

- Castor oil given at term appears to be effective in promoting labor, with a consistent side effect of maternal nausea; whether other outcomes of interest are affected is unclear.

- Manual nipple stimulation at term may promote labor; effectiveness may be dependent on the protocol used and patient ability to adhere to the protocol. Currently available data are insufficient to draw conclusions.

- Data on the effectiveness of electrical breast stimulation as a method for inducing labor in prolonged gestation are inconclusive because of small sample size and a low proportion of subjects induced for an indication of prolonged pregnancy.

- Data on the safety and effectiveness of relaxin are limited and no conclusions can be drawn.

- Sweeping of the membranes at or near term is effective in promoting labor and reducing the incidence of induction for prolonged gestation.

- In general, there is a tradeoff between the effectiveness of induction agents when effectiveness is defined in terms of achieving delivery and shortening the time to delivery on the one hand, and risks of uterine tachysystole, hyperstimulation, and potential fetal compromise on the other. In increasing order of effectiveness, slow-dose oxytocin is followed by fast-dose oxytocin; PGE_2 appears more effective than oxytocin, and misoprostol is more effective than PGE_2. The heterogeneity of the patient populations in the published literature prohibit definitive conclusions about the benefits and risks of these agents in the setting of induction of labor in prolonged pregnancy, either for women induced electively or for women with abnormal fetal surveillance.

- Mifepristone (RU-486) is consistently effective in reducing the time to labor and the time to delivery in women after 41 weeks. However, all three published trials reported nonsignificant trends towards higher rates of intermediate markers of fetal compromise, including abnormal fetal heart rate tracings and low Apgar scores.

- Data on costs are insufficient to allow conclusions about cost-effectiveness.

Question 4: Are the epidemiology and outcomes of prolonged pregnancy different for women in different ethnic groups, different socioeconomic groups, or in adolescent women?

Approach

We approached this question in two ways. First, in all the articles we reviewed, we searched for data on differences in either the epidemiology or outcomes of prolonged pregnancy in different ethnic groups, different socioeconomic groups, and different age groups. Second, we reviewed published data from birth certificates (Ventura, Martin, Curtin, et al., 2000) and from the 1997 Nationwide Inpatient Sample (NIS) (Nationwide Inpatient Sample [NIS], 1997). The NIS is part of the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project (HCUP). HCUP collects discharge data from a stratified sample of approximately 20 percent of U.S. hospitals. Using ICD-9 codes, we divided all deliveries into "preterm" (644.2x), prolonged (645.x), and term (all other delivery codes). We examined differences in outcomes between coded ethnic groups (white, black, Hispanic, Asian/Pacific Islander, Native American, and "other") and by insurance status (Medicare, Medicaid, private/health maintenance organization [HMO], self-pay/no insurance, "no charge," and "other") within these categories.

Results

Racial and Ethnic Differences: Literature Review

We did not identify any articles that specifically addressed differences in the epidemiology or outcomes of prolonged pregnancy in different ethnic groups.

Racial and Ethnic Differences: Primary Data

Birth certificate data. Table 29 summarizes total births, with percentages of infants born after 40 weeks, 41 weeks, and 42 weeks, from 1998 birth certificate data reported to the National Center for Health Statistics (NCHS), by race of mother (Asian or Native American data are not available in the published report). The proportions reported were calculated from the absolute numbers provided in the NCHS report. Table 29 also illustrates the proportion of live births after 42 weeks that were low birthweight (less than 2,500 grams) or macrosomic (greater than 4,000 grams).

Taking into account the limitations of birth certificate data, there are some interesting findings:

- Live births between 40 and 42 weeks were less common for non-Hispanic black women than for non-Hispanic white women, which may be partly due to an increased risk of preterm birth among non-Hispanic blacks (17.5 percent vs. 10.2 percent in non-Hispanic whites). However, the proportion of births after 42 weeks is strikingly similar in all groups.

- The weight distribution among infants born after 42 weeks is also strikingly different between groups, with non-Hispanic black women having a two-fold increase in low birthweight infants and a substantially lower incidence of macrosomic infants.

Hospital discharge data. Table 30 shows the percentage distribution of selected discharge diagnoses in the subset of women with a primary discharge diagnosis of prolonged pregnancy, by coded ethnic group. Total raw discharges in the NIS with this diagnosis were 57,814, or 7.2 percent of the total pregnancy-related discharges. Again, black women were more likely than women in other ethnic groups to have a diagnosis of restricted fetal growth and were less likely to have a diagnosis of macrosomia than white or Hispanic women. Black women also were more likely to have diagnoses of fetal distress and oligohydramnios. Interestingly, they also were somewhat more likely to have a diagnosis of shoulder dystocia than white or Hispanic women. Asian/Pacific Islander women were more likely to have diagnoses of macrosomia but less likely to have perineal trauma of any kind. Potential explanations for this observation include a higher cesarean section rate in Asian/Pacific Islander women, differences in the pelvic floor, or dynamics of labor which make perineal trauma less likely.

Both the NIS data and birth certificate data suggest that black women are more likely to have low birthweight infants after 42 weeks than white or Hispanic women. Diagnoses such as oligohydramnios and fetal growth restriction are also more common in black women. All three of these diagnoses are consistent with declining uteroplacental function. There were a limited number of fetal deaths in the NIS data set, with racial data missing from over half.

Socioeconomic Groups: Literature Review

We did not identify any articles that specifically addressed differences in the epidemiology or outcomes of prolonged pregnancy in different socioeconomic groups.

Socioeconomic Groups: Primary Data

Table 31 shows the percentage distribution of coded discharge diagnoses by payer status of women with a diagnosis of prolonged pregnancy. Women with private or HMO insurance coverage were less likely than women with Medicaid or no insurance to have diagnoses of intrauterine growth restriction or oligohydramnios.

Age Differences: Literature Review

We did not identify any articles that specifically addressed differences in the epidemiology or outcomes of prolonged pregnancy in either adolescent women or women in their later reproductive years.

Methodological Issues

Data Quality Issues

The accuracy of the dating recorded on birth certificates is unconfirmable, at best. Therefore, it is unclear whether the observed trends in racial differences in the distribution of birthweight after 42 weeks, and the observed lack of difference in the proportion of all pregnancies that reach 42 weeks, are real or simply random error introduced by variable quality of dating.

Similarly, criteria for a diagnosis of prolonged pregnancy, as well as for many of the other diagnosis codes, may vary between hospitals. Data for racial and payer codes were missing for many of the coded complication diagnoses. If codes are not recorded systematically in some hospitals, this may result in misleading patterns.

Statistical Analysis

Because of concerns with data quality, we did not perform formal tests of significance or multivariate analyses. Given the consistent patterns for some observations seen in the two data sets, more detailed analysis of more complete data sets is warranted.

Summary

The current published literature on the epidemiology and management of prolonged pregnancy does not provide information on the potential effects of race and ethnicity, socioeconomic status, or age on the incidence and outcomes of prolonged pregnancy. Given that many of the strategies designed to minimize the risk of fetal compromise (such as frequent antepartum testing) may have different practical effects in populations with different levels of access to transportation, child care, and appropriate monitoring facilities, this lack of information is disappointing.

Review of national data from birth certificates and hospital discharges suggests that there may be differences in the clinical characteristics of prolonged pregnancy among women in different ethnic and socioeconomic groups. In spite of the multiple limitations of the data, it is striking that two different data sources both show that black women with prolonged pregnancy

are more likely to have low birthweight infants than white or Hispanic women. Black women are consistently more likely to have low birthweight infants at other gestational ages as well. Black women also are more likely to have diagnoses of intrauterine growth restriction and oligohydramnios. Women with Medicaid or no insurance are also more likely to have growth restriction and oligohydramnios. We did not explore the degree to which the effects of race might be confounded by economic status, or vice versa, primarily because of problems caused by missing data. Other potential confounders include differences in the use of ultrasound for dating and differences in the use of antepartum testing for prolonged pregnancy. These findings should be investigated further using higher quality data and appropriate epidemiological and statistical methodologies.

Chapter 4. Conclusions

In this section we summarize the main findings of the report and discuss the implications of the findings, the limitations of the current literature, the limitations of the report, and suggested strategies for using the report to develop quality improvement tools.

Summary of Findings

The major findings and conclusions for each of the four key research questions are as follows:

1. **What are the test characteristics (reliability, sensitivity, specificity, predictive values) and costs of measures used in the management of prolonged pregnancy to (a) assess risks to the fetus and mother of prolonged pregnancy, and (b) assess the likelihood of a successful induction of labor?**

Consistently, tests for the assessment of risks to the fetus have lower sensitivity than specificity but higher negative predictive values than positive predictive values. This implies that the low risk of adverse outcomes is the main "driver" of high negative predictive values, and if sensitivity and specificity do not change appreciably with gestational age, that negative predictive value—the likelihood that a fetus with a normal test will have a normal outcome—decreases with advancing gestational age. Thus, false negative results will increase with advancing gestational age.

The most sensitive tests to assess the risks to the fetus of prolonged pregnancy appear to be combinations of fetal heart rate monitoring and ultrasonographic measurement of amniotic fluid volume. Direct comparison of test results across studies is difficult because of differences in patient populations and reference standards used. Published data on costs were not available.

Both ultrasound and clinical examination can be reasonably sensitive at identifying macrosomic fetuses when macrosomia is defined as greater than 4,000 grams. However, prediction of birthweights greater than 4,500 grams, the clinically more relevant threshold, is less accurate, with sensitivity ranges from 14-99 percent. There is no evidence that early detection of macrosomic infants in prolonged pregnancy improves maternal or neonatal outcomes, and modeling studies suggest that the use of ultrasound to screen for macrosomia is not cost effective.

The components of the cervical examination used to determine the Bishop score have significant inter- and intraobserver variability. The uncertainty created by this variability affects the ability of the examination to discriminate between patients likely to have a successful induction and those likely to fail.

2. **What is the direct evidence comparing the benefits, risks, and costs of planned induction versus expectant management at various gestational ages?**

Although individual randomized trials do not show significant differences in perinatal mortality between women electively induced at specific gestational ages and women followed with antepartum testing, pooled data show a significant reduction in perinatal mortality in women electively induced after 41 weeks compared with women managed with antepartum

testing. At least 500 inductions are needed to prevent one perinatal death. Cesarean section rates do not appear to differ between electively induced and expectantly managed women, either overall or in specific subgroups. In some groups, elective induction actually decreases the overall risk of cesarean section. Other maternal and perinatal outcomes do not appear to differ between groups.

Data on patient preferences for management options are lacking. Analysis of costs in the largest trial suggested that costs were reduced with elective induction; more detailed analysis based on currently used interventions and current obstetric management is needed.

3. What are the benefits, risks, and costs of currently available interventions for the induction of labor?

The majority of studies of interventions for induction of labor involved women induced for a variety of indications at a wide range of gestational ages. Whether summary results from these groups are applicable to women with prolonged pregnancy is unclear.

Sweeping or "stripping" of the membranes at 38-40 weeks consistently promotes spontaneous labor and reduces the number of women requiring induction at 41 or 42 weeks.

Many studies of agents for induction are powered based on detecting differences in time to induction or differences in the proportion of women delivered within a predetermined period of time. Most do not have sufficient power to detect differences in categorical outcomes, such as cesarean section rates and adverse maternal or perinatal outcomes.

There is a consistent pattern of tradeoffs between efficacy of interventions for induction, especially as measured by time to induction or delivery within a predetermined period of time, and uterine hyperactivity, with possible increased risks of surrogate markers of fetal compromise, such as nonreassuring fetal heart rate tracings. Misoprostol appears most consistently to result in vaginal delivery within a predefined time period; however, it also appears most likely to result in very frequent uterine contractions, which may lead to fetal heart rate abnormalities.

Data are lacking on both medical and nonmedical costs of different intervention strategies.

4. Are the epidemiology and outcomes of prolonged pregnancy different for women in different ethnic groups, different socioeconomic groups, or in adolescent women?

We identified no published literature that showed differences among important ethnic, socioeconomic, or other subgroups.

Review of administrative data suggests that the proportion of all pregnancies extending beyond 42 weeks is similar among all racial and ethnic groups. Black women are more likely to have low birthweight infants after 42 weeks than other groups, a finding similar to observations at other gestational ages. Confirmation of these observations with more detailed data sets is needed.

Currently available literature on interventions in prolonged gestation does not address issues such as access to care or practical difficulties (for example, transportation or arranging child care) which might affect effectiveness (as opposed to efficacy) in different populations.

Research Implications

The primary research implication of our review of the literature is that much remains to be learned about the optimal management of pregnancy in women who go beyond 40 weeks gestation with otherwise normal pregnancies. It is clear that the risks of adverse outcomes increases with advancing gestational age, but the point at which this risk justifies more intensive interventions is unclear. Currently available antepartum testing strategies have good negative predictive value but poor positive predictive value. This appears to be largely due to the overall low absolute risk of adverse outcomes, since test specificity is generally better than sensitivity. The optimal test or combination of tests and the optimal timing of test initiation among women in the United States that would minimize the risk of complications associated with prolonged gestation and complications of interventions at an acceptable cost are unclear. Several interventions are available for the effective induction of labor; however, the populations studied in the published literature are heterogeneous in terms of indications for induction. Whether the benefit/risk profile of this diverse population is equivalent to that in women induced solely because of prolonged gestation, or because of abnormal antepartum testing in prolonged gestation, is unclear. Pooled results from randomized trials comparing scheduled induction and expectant management with antepartum testing show a reduced risk of perinatal mortality in women with scheduled induction after 41 weeks, with at least 500 inductions needed to prevent one death. However, the cost-effectiveness of these strategies needs to be compared using more recent data. Administrative data suggest that there are racial and ethnic differences in the epidemiology and outcomes of prolonged pregnancy; these differences need to be explored using more detailed data sets. Finally, given the complexity of decisionmaking in settings where there often are competing risks between mother and fetus, and where patients clearly have strong preferences for the process of labor and delivery, the lack of scientific data on patient preferences, quality of life, and other “subjective” measures is impressive.

Limitations of the Current Literature

Although there are a large number of randomized trials available that provide evidence addressing the key questions identified in this report, there are numerous limitations to the current literature:

- Heterogeneity of patient populations: A consistent problem with much of the literature on specific intervention agents is inclusion of women being induced for a variety of indications. Both the benefits (in terms of successful induction) and risks (in terms of fetal compromise) of induction agents might be quite different in different populations of patients. Studies either should be performed exclusively in patients with prolonged pregnancy, or subgroup analyses should be reported so that pooled estimates of efficacy in different populations can be generated.

- Appropriate endpoints: Stillbirth is, fortunately, a rare outcome even in “high-risk” populations. Most feasible studies of tests or interventions will not have sufficient power to detect differences in mortality rates. However, the clinical utility of commonly used endpoints is compromised because of inherent unreliability and susceptibility to bias (changes in fetal heart rate pattern or cervical examination), uncertainty about long-term

clinical significance (presence of meconium in amniotic fluid or Apgar scores), and the effect of variability in knowledge of preintervention test results or local practice patterns (cesarean section rates). Finally, the lack of data on patient preferences and quality-of-life measures is striking.

- Statistical issues: Even well-done studies with a priori sample size estimates often are underpowered to detect potentially clinically relevant differences in outcomes, especially when sample size estimates are based on continuous variables (such as time to delivery) and other outcomes are categorical (such as cesarean section rates). Inappropriate measures of central tendency and statistical tests are often used (for example, treating variables such as Bishop score or parity as continuous variables). This may also lead to erroneous conclusions about differences between groups.

Limitations of the Report

Literature Search

We used standard methods for identifying, reviewing, and abstracting published studies focused on the management of prolonged pregnancy. We used predefined study characteristics to identify those studies most likely to provide unbiased estimates of efficacy and test performance. We did not search the literature prior to 1980, primarily because we assumed that the lack of general availability of ultrasound for both dating and management of prolonged gestation would limit the applicability of these results to current practice. We also limited our search to articles published in English, primarily for reasons of convenience and resource constraints. It is possible that including older studies, or studies published in other languages, would have identified additional evidence that would have substantially changed our conclusions. This may be especially true for alternative or complementary therapies.

Another limitation of our exclusion criteria is that rare but severe complications of treatments may have been overlooked because they were published in case reports or small case series. Although these study designs are useful for identifying potential problems, it is difficult to quantify these risks when only numerator values are available.

Grading of Articles

We did not use one of the currently available quality scoring systems to grade the articles we reviewed. However, we believe that the rationale for each criterion we used is reasonable, and that the operational definitions are clear and reproducible. In addition, we used these grading criteria primarily to provide additional detail to other researchers. We did not use them to establish a threshold for including or excluding articles or to weight the results of a quantitative evidence synthesis such as a meta-analysis.

Other Data Sources

We used one additional data source in preparing this report, the Nationwide Inpatient Sample (NIS) (Nationwide Inpatient Sample [NIS], 1997). The NIS, like most administrative databases, is limited by a lack of clinically relevant detail. In addition, even the data recorded in these discharge abstracts were incomplete, limiting our ability to analyze them in great detail. Variability in definitions between hospitals also may lead to incorrect conclusions. The primary value of these data in the context of this report is to identify potentially important differences in outcomes between ethnic and socioeconomic groups that need to be explored further in data sets with better documentation and more complete data.

Suggested Strategies for Using this Report

The state of the currently available evidence probably does not allow for the creation of highly specific clinical guidelines or performance measures for many aspects of managing prolonged pregnancy. Consistent conclusions from the report include:

- Sweeping of the membranes consistently promotes labor. However, given the lack of data on patient preferences for undergoing this procedure or on the value of promoting labor, using performance of membrane sweeping as a quality measure is premature. However, discussion of this option with women during the late third trimester is certainly reasonable.

- Surveillance with tests that include fetal heart rate monitoring and assessment of amniotic fluid volume or elective induction both appear to be reasonable strategies beyond 41 weeks. Patients and providers should be informed that the best current evidence strongly suggests that there is a significant increase in the risk of perinatal mortality in women managed with antepartum testing compared with women who are electively induced at 41 weeks. Because this risk is small in absolute terms, and patients may have different preferences for both the outcomes and processes of labor and delivery, both options should be discussed.

- There is no evidence to justify induction of labor solely for the indication of macrosomia (defined as estimated fetal weight greater than 4,000 grams) in prolonged pregnancy.

Chapter 5. Future Research

According to national birth certificate data, almost 18 percent of pregnancies (702,000 women) in the United States extend beyond 41 weeks, and over 7 percent (288,000 women) extend beyond 42 weeks (Ventura, Martin, Curtin, et al., 2000). Better data on optimal management of these women would have significant public health benefit.

Estimation of Risks Associated with Prolonged Gestation

Perinatal Mortality

The most precise data available come from the United Kingdom. Estimates in U.S. populations, preferably with the ability to control for the presence of other risk factors for mortality and the use of antepartum testing, are needed. Potential studies include:

- Detailed analysis of U.S. birth certificate data.

- Detailed analysis of U.S. hospital discharge data, although this will necessarily miss deliveries performed outside the hospital, such as those performed at freestanding birth centers and home births.

- Detailed analysis of administrative or computerized clinical data from large provider organizations, such as health maintenance organizations.

Because of the inherent limitations of these data sources, validation with detailed clinical records ultimately will be needed to systematically determine and describe causes of death. These data also would allow determination of the impact of various methods of dating pregnancy on perinatal mortality.

Perinatal Morbidity

Similar methods need to be applied to estimations of the risks of perinatal morbidity:

- Careful attention should be given to case definitions; again, validation of the accuracy of administrative data is needed.

- We did not identify any recent publications providing followup data on infants born after prolonged gestation. Ultimately, long-term outcomes are most important, and better data on the long-term consequences of various management strategies are needed.

Maternal Morbidity

- Again, better estimation of the risks, given current obstetric practice, is needed.

- Recently, attention has been drawn to the risks of long-term maternal consequences of labor and delivery, especially pelvic floor dysfunction. It is unclear if any of the management strategies used for prolonged pregnancy have any impact on the risks of subsequent development of pelvic floor dysfunction.

Testing Methods

Because many outcomes associated with prolonged gestation are rare, evaluations of individual tests and testing strategies will always be either limited in power or forced to rely on surrogate measures. Further research is needed on:

- Identification of surrogate measures of fetal compromise that are less susceptible to bias or observer variation.

- Study designs that could eliminate or substantially reduce the potential for verification bias because of clinician knowledge of antepartum test results.

- The optimal timing of antepartum testing.

Data on currently available tests strongly suggest that test specificity is much better than test sensitivity. In order for expectant management to compare more favorably to elective induction, research into new testing strategies should focus on improving the negative predictive value of tests by improving test sensitivity.

In addition, detailed data are needed on the medical and nonmedical costs associated with specific tests and testing strategies.

Planned Induction versus Expectant Management

Based on the available trial data, planned induction after 41 weeks appears to reduce the risk of perinatal mortality at lower cost and at no risk of increased cesarean section rates compared with expectant management. The strongest and largest trial was completed a decade ago. Whether these conclusions are still valid given current management strategies and interventions (such as misoprostol) is unclear. It also is unclear whether the extra knowledge to be gained by yet another large trial justifies the costs of such a trial. The following points should be considered:

- Decision analysis and cost-effectiveness analysis may help quantify our current degree of uncertainty. In order to be useful, modeling will require more precise data on risks, test characteristics, the effectiveness of induction, and costs in the specific population of interest. Some of these data could be provided by the research agenda discussed above. Decision and cost-effectiveness analyses will also need to consider subtle issues such as the potential

effects of increased induction rates on staffing needs for labor-and-delivery and postpartum units.

- Again, data on patient preferences for both outcomes and process are needed. For some women, the degree of certainty provided by a scheduled induction may be preferable to repeated visits for antepartum testing and uncertainty about when labor may begin. For other women, the desire to minimize intervention in the pregnancy may take precedence. How these preferences interact with patients' attitudes and preferences about risks to both themselves and their babies is an unexplored area of research with substantial implications for individual patients, clinicians, and policymakers.

Interventions for Induction

- Despite a number of randomized trials of methods for inducing labor, our ability to draw conclusions about the efficacy of various agents in women with prolonged pregnancy is limited because of the diversity of indications for induction and the diversity of gestational ages in these trials. Data on outcomes specific to the two groups of interest—women induced electively at a specific gestational age and women with prolonged pregnancy induced because of abnormal fetal heart rate testing—are needed. These data could be obtained either by performing a meta-analysis using pooled data from previous, ongoing, or future trials in these specific subgroups or by performing trials limited to these two groups.

- Sample size estimates for trials should be based on clinically relevant outcomes. Although time from beginning of induction to delivery is an important resource outcome, there are no data available on how women value this outcome compared with others. When sample size estimation is based on time-related variables, power to detect clinically relevant differences in other outcomes is diminished.

- Use of primary outcomes limited by inherent lack of reliability, such as Bishop score or abnormal fetal heart rate tracings, should be avoided. If used as secondary outcomes, consideration should be given when feasible to the use of research techniques designed to minimize the effects of observer variation, such as review by blinded outside experts (an approach often used in trials where data sources such as electrocardiograms, radiology films, or pathology slides are required).

- Patient preferences and quality-of-life measures, using standard techniques and methods for measuring these attributes, should be included in all studies. Attention should be focused not only on patient preferences for outcomes, but on process as well. All women value a healthy baby, but there may be strong preferences for the way in which this outcome is achieved.

- Detailed data are needed on medical and nonmedical costs associated with different interventions for the induction of labor in prolonged gestation and for promoting labor in women at term.

- Given that from some perspectives elective induction of labor may be preferable to expectant management, research on establishing reliable estimates of the relative safety, effectiveness, and costs of available induction agents in this particular patient population should be a high priority.

Special Populations

Preliminary analysis of administrative data suggests that additional research into possible differences in the epidemiology and outcomes of prolonged pregnancy in different ethnic and socioeconomic groups is warranted:

- Confirmation of the lack of ethnic differences in the proportion of pregnancies extending beyond 42 weeks—despite higher rates of preterm birth in black women—using data sources where confirmation of gestational age is available, would be important.

- Confirmation of the higher rate of low birthweight and other diagnoses consistent with uteroplacental insufficiency in black women with prolonged gestation is needed. If confirmed, clinical, epidemiological, basic science, and genetic studies might provide insight into the causes of this association.

- Further exploration of the potential interaction of ethnicity and economic status is needed.

References

Acker DB, Sachs BP, Friedman EA. Risk factors for shoulder dystocia. Obstet Gynecol 1985;66(6):762-8.

ACOG. Routine ultrasound in low-risk pregnancy. 1997.

ACOG. Fetal macrosomia. ACOG Practice Bulletin No. 22. 2000.

Alfirevic Z, Howarth G, Gaussmann A. Oral misoprostol for induction of labor (Cochrane Review). In: The Cochrane Library, Issue 2, 2000. Oxford: Update Software.

Alfirevic Z, Luckas M, Walkinshaw SA, et al. A randomised comparison between amniotic fluid index and maximum pool depth in the monitoring of post-term pregnancy. Br J Obstet Gynaecol 1997;104(2):207-11.

Alfirevic Z, Walkinshaw SA. A randomised controlled trial of simple compared with complex antenatal fetal monitoring after 42 weeks of gestation. Br J Obstet Gynaecol 1995;102(8):638-43.

Allott HA, Palmer CR. Sweeping the membranes: a valid procedure in stimulating the onset of labour? Br J Obstet Gynaecol 1993:100(10):898-903.

Anonymous. ACOG practice patterns. Management of postterm pregnancy. Number 6, October 1997. American College of Obstetricians and Gynecologists. Int J Gynaecol Obstet 1997;60(1):86-91.

Arabin B, Snyjders R, Mohnhaupt A, et al. Evaluation of the fetal assessment score in pregnancies at risk for intrauterine hypoxia. Am J Obstet Gynecol 1993;169(3):549-54.

Arias F. Predictability of complications associated with prolongation of pregnancy. Obstet Gynecol 1987;70(1):101-6.

Atad J, Hallak M, Auslender R, et al. A randomized comparison of prostaglandin E2, oxytocin, and the double-balloon device in inducing labor. Obstet Gynecol 1996;87(2):223-7.

Augensen K, Bergsjø P, Eikeland T, et al. Randomised comparison of early versus late induction of labour in post-term pregnancy. BMJ 1987;294(6581):1192-5.

Ayres-de-Campos D, Bernardes J, Costa-Pereira A, et al. Inconsistencies in classification by experts of cardiotocograms and subsequent clinical decision. Br J Obstet Gynaecol 1999;106(12):1307-10.

Bastian H, Keirse MJ, Lancaster PA. Perinatal death associated with planned home birth in Australia: population based study. BMJ 1998;317(7155):384-8.

Battaglia C, Larocca E, Lanzani A, et al. Doppler velocimetry in prolonged pregnancy. Obstet Gynecol 1991;77(2):213-6.

Bell RJ, Permezel M, MacLennan A, et al. A randomized, double-blind, placebo-controlled trial of the safety of vaginal recombinant human relaxin for cervical ripening. Obstet Gynecol 1993;82(3):328-33.

Bennett BB. Uterine rupture during induction of labor at term with intravaginal misoprostol. Obstet Gynecol 1997;89(5 Pt 2):832-3.

Berghella V, Rogers RA, Lescale K. Stripping of membranes as a safe method to reduce prolonged pregnancies. Obstet Gynecol 1996;87(6):927-31.

Bergsjø P, Huang GD, Yu SQ, et al. Comparison of induced versus non-induced labor in post-term pregnancy. A randomized prospective study. Acta Obstet Gynecol Scand 1989;68(8):683-7.

Bernardes J, Costa-Pereira A, Ayres-de-Campos D, et al. Evaluation of interobserver agreement of cardiotocograms. Int J Gynaecol Obstet 1997;57(1):33-7.

Bishop EH. Pelvic scoring for elective induction. Obstet Gynecol 1964;24:266-8.

Blanchette HA, Nayak S, Erasmus S. Comparison of the safety and efficacy of intravaginal misoprostol (prostaglandin E1) with those of dinoprostone (prostaglandin E2) for cervical ripening and induction of labor in a community hospital. Am J Obstet Gynecol 1999;180(6 Pt 1):1551-9.

Bochner CJ, Medearis AL, Ross MG, et al. The role of antepartum testing in the management of postterm pregnancies with heavy meconium in early labor. Obstet Gynecol 1987;69(6):903-7.

Bochner CJ, Williams J 3d, Castro L, et al. The efficacy of starting postterm antenatal testing at 41 weeks as compared with 42 weeks of gestational age. Am J Obstet Gynecol 1988;159(3):550-4.

Boulvain M, Fraser WD, Marcoux S, et al. Does sweeping of the membranes reduce the need for formal induction of labour? A randomised controlled trial. Br J Obstet Gynaecol 1998;105(1):34-40.

Boulvain M, Irion O. Stripping/sweeping the membranes for inducing labour or preventing post-term pregnancy (Cochrane Review). In: The Cochrane Library, Issue 2, 2001. Oxford: Update Software.

Brar HS, Horenstein J, Medearis AL, et al. Cerebral, umbilical, and uterine resistance using Doppler velocimetry in postterm pregnancy. J Ultrasound Med 1989;8(4):187-91.

Brennand JE, Calder AA, Leitch CR, et al. Recombinant human relaxin as a cervical ripening agent. Br J Obstet Gynaecol 1997;104(7):775-80.

Brown VA, Sawers RS, Parsons RJ, et al. The value of antenatal cardiotocography in the management of high-risk pregnancy: a randomized controlled trial. Br J Obstet Gynaecol 1982;89(9):716-22.

Buser D, Mora G, Arias F. A randomized comparison between misoprostol and dinoprostone for cervical ripening and labor induction in patients with unfavorable cervices. Obstet Gynecol 1997;89(4):581-5.

Buttino LT, Garite TJ. Intracervical prostaglandin in postdate pregnancy. A randomized trial. J Reprod Med 1990;35(2):155-8.

Cammu H, Haitsma V. Sweeping of the membranes at 39 weeks in nulliparous women: a randomised controlled trial. Br J Obstet Gynaecol 1998;105(1):41-4.

Campbell MK, Ostbye T, Irgens LM. Post-term birth: Risk factors and outcomes in a 10-year cohort of Norwegian births. Obstet Gynecol 1997;89(4):543-8.

Cardozo L, Fysh J, Pearce JM. Prolonged pregnancy: the management debate. BMJ 1986;293(6554):1059-63.

Chang CH, Chang FM. Randomized comparison of misoprostol and dinoprostone for preinduction cervical ripening and labor induction. J Formos Med Assoc 1997;96(5):366-9.

Chatterjee MS, Ramchandran K, Ferlita J, et al. Prostaglandin E2 (PGE2) vaginal gel for cervical ripening. Eur J Obstet Gynecol Reprod Biol 1991;38(3):197-202.

Chauhan SP, Sullivan CA, Lutton TC, et al. Parous patients' estimate of birth weight in postterm pregnancy. J Perinatol 1995;15(3):192-4.

Chauhan SP, Sullivan CA, Magann EF, et al. Estimate of birthweight among post-term pregnancy: Clinical versus sonographic. J Matern Fetal Med 1994;3(5):208-11.

Chayen B, Tejani N, Verma U. Induction of labor with an electric breast pump. J Reprod Med 1986;31(2):116-8.

Chervenak JL, Divon MY, Hirsch J, et al. Macrosomia in the postdate pregnancy: is routine ultrasonographic screening indicated? Am J Obstet Gynecol 1989;161(3):753-6.

Chuck FJ, Huffaker BJ. Labor induction with intravaginal misoprostol versus intracervical prostaglandin E2 gel (Prepidil gel): randomized comparison. Am J Obstet Gynecol 1995;173(4):1137-42.

Cotzias CS, Paterson-Brown S, Fisk NM. Prospective risk of unexplained stillbirth in singleton pregnancies at term: population based analysis. BMJ 1999;319(7205):287-8.

Crane J, Bennett K, Young D, et al. The effectiveness of sweeping membranes at term: a randomized trial. Obstet Gynecol 1997;89(4):586-90.

Crowley P. Interventions for preventing or improving the outcome of delivery at or beyond term (Cochrane Review). In: The Cochrane Library, Issue 2, 2000. Oxford: Update Software.

Crowley P, O'Herlihy C, Boylan P. The value of ultrasound measurement of amniotic fluid volume in the management of prolonged pregnancies. Br J Obstet Gynaecol 1984;91(5):444-8.

Curtis PD, Matthews TG, Clarke TA, et al. Neonatal seizures: the Dublin Collaborative Study. Arch Dis Child 1988;63(9):1065-8.

Devoe LD, Sholl JS. Postdates pregnancy. Assessment of fetal risk and obstetric management. J

Reprod Med 1983;28(9):576-80.

Divon MY, Haglund B, Nisell H, et al. Fetal and neonatal mortality in the postterm pregnancy: the impact of gestational age and fetal growth restriction. Am J Obstet Gynecol 1998;178(4):726-31.

Doany W, McCarty J. Outpatient management of the uncomplicated postdate pregnancy with intravaginal prostaglandin E2 gel and membrane stripping. J Matern Fetal Med 1997;6(2):71-8.

Donker DK, van Geijn HP, Hasman A. Interobserver variation in the assessment of fetal heart rate recordings. Eur J Obstet Gynecol Reprod Biol 1993;52(1):21-8.

Dyson DC, Miller PD, Armstrong MA. Management of prolonged pregnancy: induction of labor versus antepartum fetal testing. Am J Obstet Gynecol 1987;156(4):928-34.

Eden RD, Gergely RZ, Schifrin BS, et al. Comparison of antepartum testing schemes for the management of the postdate pregnancy. Am J Obstet Gynecol 1982;144(6):683-92.

Eden RD, Seifert LS, Winegar A, et al. Perinatal characteristics of uncomplicated postdate pregnancies. Obstet Gynecol 1987;69(3 Pt 1):296-9.

Egarter C, Kofler E, Fitz R, et al. Is induction of labor indicated in prolonged pregnancy? Results of a prospective randomised trial. Gynecol Obstet Invest 1989;27(1):6-9.

El-Torkey M, Grant JM. Sweeping of the membranes is an effective method of induction of labour in prolonged pregnancy: a report of a randomized trial. Br J Obstet Gynaecol 1992;99(6):455-8.

Elliott CL, Brennand JE, Calder AA. The effects of mifepristone on cervical ripening and labor induction in primigravidae. Obstet Gynecol 1998;92(5):804-9.

Elliott JP, Flaherty JF. The use of breast stimulation to prevent postdate pregnancy. Am J Obstet Gynecol 1984;149(6):628-32.

Escudero F, Contreras H. A comparative trial of labor induction with misoprostol versus oxytocin. Int J Gynaecol Obstet 1997;57(2):139-43.

Evans MI, Dougan MB, Moawad AH, et al. Ripening of the human cervix with porcine ovarian relaxin. Am J Obstet Gynecol 1983;147(4):410-4.

Farah LA, Sanchez-Ramos L, Rosa C, et al. Randomized trial of two doses of the prostaglandin E1 analog misoprostol for labor induction. Am J Obstet Gynecol 1997;177(2):364-9.

Farmakides G, Schulman H, Winter D, et al. Prenatal surveillance using nonstress testing and Doppler velocimetry. Obstet Gynecol 1988;71(2):184-7.

Fleischer A, Schulman H, Farmakides G, et al. Antepartum nonstress test and the postmature pregnancy. Obstet Gynecol 1985;66(1):80-3.

Fletcher H, Mitchell S, Frederick J, et al. Intravaginal misoprostol versus dinoprostone as cervical ripening and labor-inducing agents. Obstet Gynecol 1994;83(2):244-7.

Fletcher HM, Mitchell S, Simeon D, et al. Intravaginal misoprostol as a cervical ripening agent. Br J Obstet Gynaecol 1993;100(7):641-4.

Flynn AM, Kelly J, Mansfield H, et al. A randomized controlled trial of non-stress antepartum cardiotocography. Br J Obstet Gynaecol 1982;89(6):427-33.

Froen JF, Arnestad M, Frey K, et al. Risk factors for sudden intrauterine unexplained death: epidemiologic characteristics of singleton cases in Oslo, Norway, 1986-1995. American Journal of Obstetrics & Gynecology 2001;184(4):694-702.

Frydman R, Lelaidier C, Baton-Saint-Mleux C, et al. Labor induction in women at term with mifepristone (RU 486): a double-blind, randomized, placebo-controlled study. Obstet Gynecol 1992;80(6):972-5.

Garry D, Figueroa R, Guillaume J, et al. Use of castor oil in pregnancies at term. Altern Ther Health Med 2000;6(1):77-9.

Giacalone PL, Targosz V, Laffargue F, et al. Cervical ripening with mifepristone before labor induction: A randomized study. Obstet Gynecol 1998;92(4 I):487-92.

Gilby JR, Williams MC, Spellacy WN. Fetal abdominal circumference measurements of 35 and 38 cm as predictors of macrosomia. J Reprod Med 2000;45:936-8.

Gilson GJ, O'Brien ME, Vera RW, et al. Prolonged pregnancy and the biophysical profile. A birthing center perspective. J Nurse Midwifery 1988;33(4):171-7.

Gjessing HK, Skjaerven R, Wilcox AJ. Errors in gestational age: evidence of bleeding early in pregnancy. Am J Public Health 1999;89(2):213-8.

Goeree R, Hannah M, Hewson S. Cost-effectiveness of induction of labour versus serial antenatal monitoring in the Canadian Multicentre Postterm Pregnancy Trial. CMAJ 1995;152(9):1445-50.

Goldberg J, Newman RB, Rust PF. Interobserver reliability of digital and endovaginal ultrasonographic cervical length measurements. Am J Obstet Gynecol 1997;177(4):853-8.

Gonen O, Rosen DJ, Dolfin Z, et al. Induction of labor versus expectant management in macrosomia: a randomized study. Obstet Gynecol 1997;89(6):913-7.

Gottschall DS, Borgida AF, Mihalek JJ, et al. A randomized clinical trial comparing misoprostol with prostaglandin E2 gel for preinduction cervical ripening. Am J Obstet Gynecol 1997;177(5):1067-70.

Graves GR, Baskett TF, Gray JH, et al. The effect of vaginal administration of various doses of prostaglandin E2 gel on cervical ripening and induction of labor. Am J Obstet Gynecol 1985;151(2):178-81.

Grünberger W, Spona J. The effect of pericervical PGEinf 2 instillation on levels of maternal serum 13,14 dihydro-15-keto-PGF(2alpha) and progesterone. Arch Gynecol 1986;239(2):93-9.

Gupta R, Vasishta K, Sawhney H, et al. Safety and efficacy of stripping of membranes at term. Int J Gynaecol Obstet 1998;60(2):115-21.

Hann L, McArdle C, Sachs B. Sonographic biophysical profile in the postdate pregnancy. J Ultrasound Med 1987;6(4):191-5.

Hannah ME, Hannah WJ, Hellmann J, et al. Induction of labor as compared with serial antenatal monitoring in post-term pregnancy. A randomized controlled trial. The Canadian Multicenter Post-term Pregnancy Trial Group [published erratum appears in N Engl J Med 1992 Jul 30;327(5):368]. N Engl J Med 1992;326(24):1587-92.

Hannah ME, Huh C, Hewson SA, et al. Postterm pregnancy: putting the merits of a policy of induction of labor into perspective. Birth 1996;23(1):13-9.

Harris BA, Huddleston JF, Sutliff G, et al. The unfavorable cervix in prolonged pregnancy. Obstet Gynecol 1983;62(2):171-4.

Hedén L, Ingemarsson I, Ahlström H, et al. Induction of labor versus conservative management in prolonged pregnancy: controlled study. Int J Feto-Matern Med 1991;4(4):231-6.

Henriksen TB, Wilcox AJ, Hedegaard M, et al. Bias in studies of preterm and postterm delivery due to ultrasound assessment of gestational age. Epidemiology 1995;6(5):533-7.

Herabutya Y, Prasertsawat P, Pokpirom J. A comparison of intravaginal misoprostol and intracervical prostaglandin E2 gel for ripening of unfavorable cervix and labor induction. J Obstet Gynaecol Res 1997;23(4):369-74.

Herabutya Y, Prasertsawat PO, Tongyai T, et al. Prolonged pregnancy: the management dilemma. Int J Gynaecol Obstet 1992;37(4):253-8.

Hilder L, Costeloe K, Thilaganathan B. Prolonged pregnancy: evaluating gestation-specific risks of fetal and infant mortality. Br J Obstet Gynaecol 1998;105(2):169-73.

Hofmeyr GJ, Gulmezoglu AM. Vaginal misoprostol for cervical ripening and labour induction in late pregnancy (Cochrane Review). In: The Cochrane Library, Issue 1, 2001. Oxford: Update Software.

Holcomb WL, Smeltzer JS. Cervical effacement: variation in belief among clinicians. Obstet Gynecol 1991;78(1):43-5.

Howarth GR, Funk M, Steytler P, et al. A randomised controlled trial comparing vaginally administered misoprostol to vaginal dinoprostone gel in labour induction. J Obstet Gynaecol 1996;16:474-8.

Idrisa A, Obisesan KA, Adeleye JA. Fetal membrane sweeping for stimulation of labour in prolonged pregnancy: a controlled study. J Obstet Gynaecol 1993;13(4):235-7.

Imai M, Tani A, Saito M, et al. Significance of fetal fibronectin and cytokine measurement in the cervicovaginal secretions of women at term in predicting term labor and post-term pregnancy. Eur J Obstet Gynecol Reprod Biol 2001;97:53-8.

Jazayeri A, Heffron JA, Phillips R, et al. Macrosomia prediction using ultrasound fetal abdominal circumference of 35 centimeters or more. Obstet Gynecol 1999;93(4):523-6.

Johanson RB, Menon BKV. Vacuum extraction vs forcepts for assisted vaginal delivery (Cochrane Review). In: The Cochrane Library, Issue 2, 2001. Oxford: Update Software.

Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. BMJ 2001;323:42-6.

Jüni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282(11):1054-60.

Kadanali S, Küçüközkan T, Zor N, et al. Comparison of labor induction with misoprostol vs. oxytocin/prostaglandin E2 in term pregnancy. Int J Gynaecol Obstet 1996;55(2):99-104.

Kadar N, Tapp A, Wong A. The influence of nipple stimulation at term on the duration of pregnancy. J Perinatol 1990;10(2):164-6.

Katz Z, Yemini M, Lancet M, et al. Non-aggressive management of post-date pregnancies. Eur J Obstet Gynecol Reprod Biol 1983;15(2):71-9.

Kelly AJ, Kavanagh J, Thomas J. Castor oil, bath and/or enema for cervical priming and induction of labour (Cochrane Review). In: The Cochrane Library, Issue 2, 2001. Oxford: Update Software.

Kemp B, Winkler M, Rath W. Induction of labor by prostaglandin E2 in relation to the Bishop score. Int J Gynecol Obstet 2000;71:13-7.

Keng MTY, Eng BTS. Oral contraceptives postdatism as a complication. Med J Malaysia 1982;37(4):338-43.

Kidd LC, Patel NB, Smith R. Non-stress antenatal cardiotocography—a prospective randomized clinical trial. Br J Obstet Gynaecol 1985;92(11):1156-9.

Knox GE, Huddleston JF, Flowers CE. Management of prolonged pregnancy: results of a prospective randomized trial. Am J Obstet Gynecol 1979;134(4):376-84.

Kramer RL, Gilson GJ, Morrison DS, et al. A randomized trial of misoprostol and oxytocin for induction of labor: safety and efficacy. Obstet Gynecol 1997;89(3):387-91.

Lee HY. A randomised double-blind study of vaginal misoprostol vs dinoprostone for cervical ripening and labour induction in prolonged pregnancy. Singapore Med J 1997;38(7):292-4.

Leveno KJ, Quirk JG, Cunningham FG, et al. Prolonged pregnancy. I. Observations concerning the causes of fetal distress. Am J Obstet Gynecol 1984;150(5 Pt 1):465-73.

Lidegaard O, Bottcher LM, Weber T. Description, evaluation and clinical decision making according to various fetal heart rate patterns. Inter-observer and regional variability. Acta Obstet Gynecol Scand 1992;71(1):48-53.

Lien JM, Morgan MA, Garite TJ, et al. Antepartum cervical ripening: applying prostaglandin E2 gel in conjunction with scheduled nonstress tests in postdate pregnancies. Am J Obstet Gynecol 1998;179(2):453-8.

Lumley J, Lester A, Anderson I, et al. A randomized trial of weekly cardiotocography in high-risk obstetric patients. Br J Obstet Gynaecol 1983;90(11):1018-26.

Lydon-Rochelle M, Holt VL, Easterling TR, et al. Risk of uterine rupture during labor among women with a prior cesarean delivery. N Engl J Med 2001;345(1):3-8.

MacKenzie IZ, Burns E. Randomised trial of one versus two doses of prostaglandin E2 for induction of labour: 1. Clinical outcome. Br J Obstet Gynaecol 1997;104(9):1062-7.

Magann EF, Chauhan SP, Nevils BG, et al. Management of pregnancies beyond forty-one weeks' gestation with an unfavorable cervix. Am J Obstet Gynecol 1998;178(6):1279-87.

Magann EF, McNamara MF, Whitworth NS, et al. Can we decrease postdatism in women with an unfavorable cervix and a negative fetal fibronectin test result at term by serial membrane sweeping? Am J Obstet Gynecol 1998;179(4):890-4.

Mahmood TA. A prospective comparative study on the use of prostaglandin E2 gel (2 mg) and prostaglandin E2 tablet (3 mg) for the induction of labour in primigravid women with unfavorable cervices. Eur J Obstet Gynecol Reprod Biol 1989;33(2):169-75.

Mannino F. Neonatal complications of postterm gestation. J Reprod Med Obstet Gynecol 1988;33(3):271-6.

Martin JN, Sessums JK, Howard P, et al. Alternative approaches to the management of gravidas with

prolonged-postterm-postdate pregnancies. J Miss State Med Assoc 1989;30(4):105-11.

McColgin SW, Hampton HL, McCaul JF, et al. Stripping membranes at term: can it safely reduce the incidence of post-term pregnancies? Obstet Gynecol 1990:76(4):678-80.

McColgin SW, Patrissi GA, Morrison JC. Stripping the fetal membranes at term. Is the procedure safe and efficacious? Journal of Reproductive Medicine 1990;35(8):811-4.

Mehl-Madrona L, Madrona MM. Physician- and midwife-attended home births. Effects of breech, twin, and post-dates outcome data on mortality rates. J Nurse-Midwifery 1997;42(2):91-8.

Misra M, Vavre S. Labour induction with intracervical prostaglandin Einf 2 gel and intravenous oxytocin in women with a very unfavourable cervix. Aust N Z J Obstet Gynaecol 1994;34(5):511-5.

Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. Internat J Tech Assess Health Care 1996;12(2):195-208.

Monaghan J, O'Herlihy C, Boylan P. Ultrasound placental grading and amniotic fluid quantitation in prolonged pregnancy. Obstet Gynecol 1987;70(3 Pt 1):349-52.

Montan S, Malcus P. Amniotic fluid index in prolonged pregnancy: A cohort study. J Matern Fetal Invest 1995;5(1):4-7.

Mouw RJ, Egberts J, Kragt H, et al. Cervicovaginal fetal fibronectin concentrations: predictive value of impending birth in postterm pregnancies. Eur J Obstet Gynecol Reprod Biol 1998;80(1):67-70.

Mundle WR, Young DC. Vaginal misoprostol for induction of labor: a randomized controlled trial. Obstet Gynecol 1996;88(4 Pt 1):521-5.

National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. A clinical trial of induction of labor versus expectant management in postterm pregnancy. The National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. Am J Obstet Gynecol 1994;170(3):716-23.

Nationwide Inpatient Sample [NIS]. [electronic database]. Release 6; 1997 data. Agency for Healthcare Research and Quality: Rockville, MD. Available from: http://www.ntis.gov/fcpc/cpn8834.htm. 1997.

Neilson JP. Mifepristone for induction of labour (Cochrane Review). In: The Cochrane Library, Issue 3, 2001. Oxford: Update Software.

Nocon JJ, McKenzie DK, Thomas LJ, et al. Shoulder dystocia: an analysis of risks and obstetric maneuvers. Am J Obstet Gynecol 1993;168(6 Pt 1):1732-7.

O'Brien JM, Mercer BM, Cleary NT, et al. Efficacy of outpatient induction with low-dose intravaginal prostaglandin E2: a randomized, double-blind, placebo-controlled trial. Am J Obstet Gynecol 1995;173(6):1855-9.

O'Reilly-Green CP, Divon MY. Predictive value of amniotic fluid index for oligohydramnios in patients with prolonged pregnancies. J Matern Fetal Med 1996;5(4):218-26.

O'Reilly-Green CP, Divon MY. Receiver operating characteristic curves of sonographic estimated fetal weight for prediction of macrosomia in prolonged pregnancies. Ultrasound Obstet Gynecol 1997;9(6):403-8.

Ohel G, Rahav D, Rothbart H, et al. Randomised trial of outpatient induction of labor with vaginal PGE2 at 40-41 weeks of gestation versus expectant management. Arch Gynecol Obstet 1996;258(3):109-12.

Papageorgiou I, Tsionou C, Minaretzis D, et al. Labor characteristics of uncomplicated prolonged pregnancies after induction with intracervical prostaglandin E2 gel versus intravenous oxytocin. Gynecol Obstet Invest 1992;34(2):92-6.

Pattison N, McCowan L. Cardiotocography for antepartum fetal assessment (Cochrane Review). In: The Cochrane Library, Issue 1, 2001. Oxford: Update Software.

Pearce JM, Cardozo C. Prolonged pregnancy: the management debate. BMJ 1988;297:715-7.

Phelan JP, Platt LD, Yeh SY, et al. The role of ultrasound assessment of amniotic fluid volume in the management of the postdate pregnancy. Am J Obstet Gynecol 1985;151(3):304-8.

Phelan JP, Platt LD, Yeh SY, et al. Continuing role

of the nonstress test in the management of postdates pregnancy. Obstet Gynecol 1984;64(5):624-8.

Phelps JY, Higby K, Smyth MH, et al. Accuracy and intraobserver variability of simulated cervical dilatation measurements. Am J Obstet Gynecol 1995;173(3 Pt 1):942-5.

Plaut MM, Schwartz ML, Lubarsky SL. Uterine rupture associated with the use of misoprostol in the gravid patient with a previous cesarean section. Am J Obstet Gynecol 1999;180(6 Pt 1):1535-42.

Pollack RN, Hauer-Pollack G, Divon MY. Macrosomia in postdates pregnancies: the accuracy of routine ultrasonographic screening. Am J Obstet Gynecol 1992;167(1):7-11.

Ramrekersingh-White P, Farkas AG, Chard T, et al. Self-selected expectant management of post dates pregnancy including the use of Doppler ultrasound. J Obstet Gynaecol 1993;13(1):16-9.

Rayburn W, Gosen R, Ramadei C, et al. Outpatient cervical ripening with prostaglandin E2 gel in uncomplicated postdate pregnancies. Am J Obstet Gynecol 1988;158(6 Pt 1):1417-23.

Rayburn WF, Motley ME, Stempel LE, et al. Antepartum prediction of the postmature infant. Obstet Gynecol 1982;60(2):148-53.

Roberts LJ, Young KR. The management of prolonged pregnancy—an analysis of women's attitudes before and after term. Br J Obstet Gynaecol 1991;98(11):1102-6.

Rouse DJ, Owen J, Goldenberg RL, et al. The effectiveness and costs of elective cesarean delivery for fetal macrosomia diagnosed by ultrasound. JAMA 1996;276(18):1480-6.

Royal College of Obstetricians and Gynaecologists. Induction of labour: evidence-based clinical guideline number 9. London: Royal College of Obstetricians and Gynaecologists, 2001.

Salamalekis E, Vitoratos N, Kassanos D, et al. Sweeping of the membranes versus uterine stimulation by oxytocin in nulliparous women. A randomized controlled trial. Gynecol Obstet Invest 2000;49(4):240-3.

Sanchez-Ramos L, Kaunitz AM, Del Valle GO, et al. Labor induction with the prostaglandin E1 methyl analogue misoprostol versus oxytocin: a randomized trial. Obstet Gynecol 1993;81(3):332-6.

Sarkar PK, Duthie SJ. The clinical significance of reduced amniotic fluid index in post-term pregnancy: a retrospective study. J Obstet Gynaecol 1997;17(3):274-5.

Sarno AP, Hinderstein WN, Staiano RA. Fetal macrosomia in a military hospital: Incidence, risk factors, and outcome. Mil Med 1991;156(2):55-8.

Satin AJ, Hankins GD, Yeomans ER. A prospective study of two dosing regimens of oxytocin for the induction of labor in patients with unfavorable cervices. Am J Obstet Gynecol 1991;165(4 Pt 1):980-4.

Sawai SK, O'Brien WF, Mastrogiannis DS, et al. Patient-administered outpatient intravaginal prostaglandin E2 suppositories in post-date pregnancies: a double-blind, randomized, placebo-controlled study. Obstet Gynecol 1994;84(5):807-10.

Sawai SK, Williams MC, O'Brien WF, et al. Sequential outpatient application of intravaginal prostaglandin E2 gel in the management of postdates pregnancies. Obstet Gynecol 1991;78(1):19-23.

Schreyer P, Bar-Natan N, Sherman DJ, et al. Fetal breathing movements before oxytocin induction in prolonged prenancies. Am J Obstet Gynecol 1991;165(3):577-81.

Sciscione AC, Nguyen L, Manley J, et al. A randomized comparison of transcervical Foley catheter to intravaginal misoprostol for preinduction cervical ripening. Obstet Gynecol 2001;97:603-7.

Sherer DM, Onyeije CI, Binder D, et al. Uncomplicated baseline fetal tachycardia or bradycardia in postterm pregnancies and perinatal outcome. Am J Perinatol 1998;15(5):335-8.

Shime J, Librach CL, Gare DJ, et al. The influence of prolonged pregnancy on infant development at one and two years of age: a prospective controlled study. Am J Obstet Gynecol 1986;154(2):341-5.

Small ML, Phelan JP, Smith CV, et al. An active management approach to the postdate fetus with a reactive nonstress test and fetal heart rate decelerations. Obstet Gynecol 1987;70(4):636-40.

Smith GCS. Life-table analysis of the risk of perinatal death at term and post term in singleton pregnancies. Am J Obstet Gynecol 2001;184:489-96.

Srisomboon J, Piyamongkol W, Aiewsakul P. Comparison of intracervical and intravaginal misoprostol for cervical ripening and labour induction in patients with an unfavourable cervix. J Med Assoc Thai 1997;80(3):189-94.

Stenlund PM, Ekman G, Aedo A-R, et al. Induction of labor with mifepristone: a randomized, double-blind study versus placebo. Acta Obstet Gynecol Scand 1999;78(9):793-8.

Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283(15):2008-12.

Surbek DV, Boesiger H, Hoesli I, et al. A double-blind comparison of the safety and efficacy of intravaginal misoprostol and prostaglandin E2 to induce labor. Am J Obstet Gynecol 1997;177(5):1018-23.

Tai-Seale M, Rodwin M, Wedig G. Drive-through delivery: where are the "savings"? Med Care Res Rev 1999;56(1):30-46.

Tam W-H, Tai SMB, Rogers MS. Prediction of cervical response to prostaglandin Einf 2 using fetal fibronectin. Acta Obstet Gynecol Scand 1999;78(10):861-5.

Tongsong T, Srisomboon J. Amniotic fluid volume as a predictor of fetal distress in postterm pregnancy. Int J Gynaecol Obstet 1993;40(3):213-7.

Toppozada MK, Anwar MY, Hassan HA, et al. Oral or vaginal misoprostol for induction of labor. Int J Gynaecol Obstet 1997;56(2):135-9.

Tuffnell DJ, Bryce F, Johnson N, et al. Simulation of cervical changes in labour: reproducibility of expert assessment. Lancet 1989;2(8671):1089-90.

Varaklis K, Gumina R, Stubblefield PG. Randomized controlled trial of vaginal misoprostol and intracervical prostaglandin E2 gel for induction of labor at term. Obstet Gynecol 1995;86(4 Pt 1):541-4.

Ventura SJ, Martin JA, Curtin SC, et al. Births: Final data for 1998. National Vital Statistics Report, vol 48, no 3. Hyattsville, MD. National Center for Health Statistics, 2000.

Voss DH, Cumminsky KC, Cook VD, et al. Effects of three concentrations of intracervical prostaglandin Einf 2 gel for cervical ripening. J Matern Fetal Med 1996;5(4):186-93.

Weeks JW, Pitman T, Spinnato JA 2nd. Fetal macrosomia: does antenatal prediction affect delivery route and birth outcome? Am J Obstet Gynecol 1995;173(4):1215-9.

Weiner Z, Farmakides G, Schulman H, et al. Computerized analysis of fetal heart rate variation in postterm pregnancy: prediction of intrapartum fetal distress and fetal acidosis. Am J Obstet Gynecol 1994;171(4):1132-8.

Weiner Z, Reichler A, Zlozover M, et al. The value of Doppler ultrasonography in prolonged pregnancies. Eur J Obstet Gynecol Reprod Biol 1993;48(2):93-7.

Wing DA, Fassett MJ, Mishell DR. Mifepristone for preinduction cervical ripening beyond 41 weeks' gestation: a randomized controlled trial. Obstet Gynecol 2000;96(4):543-8.

Wing DA, Jones MM, Rahall A, et al. A comparison of misoprostol and prostaglandin E2 gel for preinduction cervical ripening and labor induction. Am J Obstet Gynecol 1995;172(6):1804-10.

Wing DA, Ortiz-Omphroy G, Paul RH. A comparison of intermittent vaginal administration of misoprostol with continuous dinoprostone for cervical ripening and labor induction. Am J Obstet Gynecol 1997;177(3):612-8.

Wing DA, Paul RH. A comparison of differing dosing regimens of vaginally administered misoprostol for preinduction cervical ripening and labor induction [published erratum appears in Am J Obstet Gynecol 1997 Jun;176(6):1423]. Am J Obstet Gynecol 1996;175(1):158-64.

Wing DA, Rahall A, Jones MM, et al. Misoprostol: an effective agent for cervical ripening and labor induction. Am J Obstet Gynecol 1995;172(6):1811-6.

Wiriyasirivaj B, Vutyavanich T, Ruangsri RA. A randomized controlled trial of membrane stripping at term to promote labor. Obstet Gynecol 1996;87(5 Pt 1):767-70.

Witter FR, Weitz CM. A randomized trial of induction at 42 weeks gestation versus expectant management for postdates pregnancies. Am J Perinatol 1987;4(3):206-11.

Witter FR, Weitz CM. Cervical examination prior to induction in postdate pregnancies. Surg Gynecol Obstet 1989;168(3):214-6.

Yudkin PL, Wood L, Redman CWG. Risk of unexplained stillbirth at different gestational ages. Lancet 1987;1:1192-4.

List of Abbreviations and Acronyms Used in the Report and Evidence Tables

Abd C	Abdominal circumference
abn	Abnormal
ACOG	American College of Obstetricians and Gynecologists
AFI	Amniotic fluid index
AFV	Amniotic fluid volume
AHRQ	Agency for Healthcare Research and Quality
APT	Antepartum testing
ARD	Atad Ripener Device
AROM	Artificial rupture of the membranes
BP	Biophysical profile
bpm	Beats per minute
BPS	Biophysical profile score
BW	birthweight
cc	Cubic centimeter(s)
CDSR	Cochrane Database of Systematic Reviews
CE	Cost-effectiveness
CI	Confidence interval
cm	Centimeter
C-section	Cesarean section
CST	Contraction stress test
CTG	Cardiotocography
DARE	Database of Abstracts of Reviews of Effectiveness
EBW	Estimated birthweight
E:C	Estrogen-to-creatinine ratio
EFW	Estimated fetal weight
FB	Fetal breathing
FBM	Fetal breathing movements
fFN	Fetal fibronectin
FHR	Fetal heart rate
FM	Fetal movement
f/u	Followup
g	Gram(s)
GP	General practitioner
HCUP	Healthcare Cost and Utilization Project
HMO	Health maintenance organization
hr	Hour(s)
IQ	Interquartile
IU	International Units(s)
IUGR	Intrauterine growth retardation
kg	Kilogram(s)
LGA	Large for gestational age
LMP	Last menstrual period
MBP	Modified biophysical profile
MFM	Maternal and family medicine
µg	Microgram(s)
mg	Milligram
min	Minute(s)
mIU	Milli-Inerantional Unit(s)
ml	Milliliter(s)
mm	Millimeter(s)
mmHg	Millimeters of mercury
MPD	Maximum pool depth
mU	Milliunit(s)
NA	Not applicable
NCHS	National Center for Health Statistics
ng	Nanogram(s)
NICHD	National Institute of Child Health and Human Development
NICU	Neonatal intensive care unit
NIS	Nationwide Inpatient Sample
nl	Normal
No.	Number
NR	Not reported
NS	Nipple stimulation
NST	Nonstress test
OB/GYN	Obstetrician/gynecologist
OCP	Oral contraceptive pill
OCT	Oxytocin challenge test
OST	Oxytocin stress test
OR	Odds ratio

PGE_2	Prostaglandin E_2 (dinoprostone)
PROM	Premature rupture of the membranes
RCOG	Royal College of Obstetricians and Gynaecologists
RCT(s)	Randomized controlled trial(s)
ROC	Receiver operating characteristic
RR	Relative risk
SD	Standard deviation
S:D	Systolic-to-diastolic ratio
sec	Second(s)
SEM	Standard error of the mean
SGA	Small for gestational age
SROM	Spontaneous rupture of the membranes
U/S	Ultrasound
UTI	Urinary tract infection
vs.	Versus
wk	Week(s)

Evidence Table 1: Studies relevant to Key Question 1

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Alfirevic, Luckas, Walkinshaw, et al., 1997	Design: RCT, randomization by sealed envelope Test(s) studied: 1) U/S measurement of amniotic fluid index (AFI) + computerized cardiotocography (CTG) using Oxford Sonicaid 8000 fetal monitor (n = 250) Protocol: If AFI < 7.3 cm (< 3rd percentile for 42-wk gestation) or if CTG abnormal (according to proprietary criteria), then labor induced. If AFI and CTG normal, then f/u visit arranged 3 days later, unless patient had reached 43 wks gestation (301 days), in which case labor induced regardless of test results. Labor induced with intravaginal prostaglandins (details NR). 2) U/S measurement of maximum pool depth (MPD) + computerized cardiotocography (CTG) using Oxford Sonicaid 8000 fetal monitor (n = 250) Protocol: If MPD < 1.8 cm (< 3rd percentile for 42-wk gestation) or if CTG abnormal (according to proprietary criteria), then labor induced. If MPD and CTG normal, then f/u visit arranged 3 days later, unless patient had reached 43 wks gestation (301 days), in which case labor induced regardless of test results.	No. of subjects at start: 500 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 500 Inclusion criteria: Uncomplicated singleton pregnancy; ≥ 40 wks gestation Exclusion criteria: Hypertension (≥ 140/95 mmHg); significant proteinuria (> 1+ on dipstick); history of antepartum hemorrhage; poor obstetric history; prior U/S suggesting IUGR Age (median, with interquartile [IQ] range): AFI + CTG: 28 (24-31); MPD + CTG: 28 (23-32) Race: NR Gestational age at entry (median, with IQ range): AFI + CTG: 290 days (289-291); MPD + CTG: 290 days (289-291) Dating criteria: 1) Certain LMP + U/S prior to 20 wks or 2) agreement within 1 wk between certain LMP and U/S after 20 wks Parity: AFI + CTG: 50% nulliparous; MPD + CTG: 50% nulliparous Bishop score: NR	1) Birthweight 2) Cord pH at delivery 3) Apgar < 7 at 5 minutes 4) Admission to NICU 5) Perinatal death 6) Cord base excess 7) Meconium 8) C-sections 9) Inductions	1) Birthweight (median, with IQ range): AFI + CTG: 3740 g (3417.5 to 3985) MPD + CTG: 3710 g (3390 to 4027.5) p = 0.89 2) Cord pH at delivery (median, with IQ range): AFI + CTG: 7.29 (7.25 to 7.34) MPD + CTG: 7.3 (7.25 to 7.34) p = 0.57 3) Apgar < 7 at 5 minutes: AFI + CTG: 5/250 (2%) MPD + CTG: 5/250 (2%) p = 1 4) Admission to NICU: AFI + CTG: 4/250 (1.6%) MPD + CTG: 4/250 (1.6%) p = 1 5) Perinatal death: AFI + CTG: 0/250 MPD + CTG: 0/250 p = 1 6) Cord base excess (median, with IQ range): AFI + CTG: -5.2 (-3.45 to -7.1) MPD + CTG: -5.4 (-3.9 to -7.2) p = 0.18 7) Meconium: AFI + CTG: 56/250 (22%) MPD + CTG: 56/250 (22%) p = 1 8) C-sections: *Overall:* AFI + CTG: 47/250 (19%) MPD + CTG: 33/250 (13%) p = 0.11	QUALITY SCORES: TESTING Reference standard: - Randomized: + Method of randomization: + Verification bias: + Test reliability/variability: + Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: + Testing protocol described: + Sample size: + Statistical tests: + MANAGEMENT Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: - Sample size estimates based on difference in C-section rates – power to detect differences in perinatal outcomes questionable.

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Labor induced with intravaginal prostaglandins (details NR). Reference standard(s): None Dates: July 1994-July 1995 Location: Liverpool, UK Setting: University hospital Type(s) of providers: General OB/GYN, MFM, midwives (nonnurse) Length of follow-up: None			*For fetal distress:* AFI + CTG: 20/250 (8%) MPD + CTG: 10/250 (4%) p = 0.09 *For failure to progress:* AFI + CTG: 25/250 (10%) MPD + CTG: 21/250 (8%) p = 0.64 *For other indications:* AFI + CTG: 2/250 (0.8%) MPD + CTG: 2/250 (0.8%) p = 1 9) Inductions: *Overall:* AFI + CTG: 87/250 (35%) MPD + CTG: 77/250 (31%) p = 0.39 *For abnormal post-term monitoring:* AFI + CTG: 37/250 (15%) MPD + CTG: 21/250 (8%) p = 0.04 *Maternal request:* AFI + CTG: 24/250 (10%) MPD + CTG: 25/250 (10%) p = 1 *43 weeks' gestation:* AFI + CTG: 17/250 (7%) MPD + CTG: 21/250 (8%) p = 0.61 *For other indications:* AFI + CTG: 9/250 (4%) MPD + CTG: 10/250 (4%) p = 1	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Alfirevic and Walkinshaw, 1995	Design: RCT, randomization by sealed envelope Test(s) studied: 1) Simple monitoring = cardiotocography (CTG) + U/S measurement of maximum pool depth (MPD) (n = 73) Protocol: If CTG abnormal (< 2 accelerations [15 bpm lasting ≥ 15 sec] in 40 min or short-term variability ≤ 5 bpm with no decelerations) or MPD abnormal (< 2.1 cm), then labor induced. If both tests normal, then f/u visit arranged 3 days later, unless patient had reached 43 wks gestation, in which case labor induced regardless of test results. Labor induced with intravaginal prostaglandins (details NR). 2) Complex monitoring = modified biophysical profile (MBP) = computerized cardiotocography (using the Oxford Sonicaid 8000 fetal monitor) + U/S measurement of amniotic fluid index (AFI) + fetal breathing movements + fetal tone + fetal gross body measurements (last 3 all monitored by U/S) (n = 72) Protocol: If AFI < 7.3 cm (< 3rd percentile for 42 wks gestation), then labor induced. If MBP total score ≤ 6 of possible 10 (each component score 0 to 2, with 2 = normal), then labor induced. If AFI	No. of subjects at start: 145 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 145 Inclusion criteria: Uncomplicated singleton pregnancy; ≥ 41 wks gestation Exclusion criteria: Hypertension (≥ 140/95 mmHg); significant proteinuria (> 1+ on dipstick); history of antepartum hemorrhage; poor obstetric history; prior U/S suggesting IUGR Age (median, with interquartile [IQ] range): Simple, 28 (25-32); complex, 29 (25-31) Race: NR Gestational age at entry: NR; gestational age ≥ 41 weeks required for entry into study Dating criteria: Certain LMP or U/S prior to 20 weeks Parity: Simple, 33% nulliparous; complex: 40% nulliparous Bishop score: NR	1) Perinatal death 2) Admission to NICU 3) Apgar score < 7 at 5 minutes 4) Cord pH at delivery 5) Meconium 6) C-sections 7) Spontaneous labor 8) Inductions 9) Normal vaginal delivery 10) Abnormal CTG intrapartum	1) Perinatal death: Simple: 0/73 Complex: 1/72 (1%) (no p-value reported) 2) Admission to NICU: Simple: 2/73 (3%) Complex: 0/72 (no p-value reported) 3) Apgar score < 7 at 5 minutes: Simple: 0/73 Complex: 1/72 (1%) (no p-value reported) 4) Cord pH at delivery (median, with IQ range): Simple: 7.31 (7.26 to 7.35) Complex: 7.29 (7.25 to 7.33) p = 0.15 5) Meconium: Simple: 14/73 (19%) Complex: 20/72 (28%) p = 0.30 6) C-sections: *Overall:* Simple: 7/73 (10%) Complex: 13/72 (18%) p = 0.22 *For fetal distress:* Simple: 6/73 (8%) Complex: 8/72 (11%) p = 0.54 *For antepartum distress:* Simple: 2/73 (3%) Complex: 0/72 (no p-value reported)	QUALITY SCORES: TESTING Reference standard: - Randomized: + Method of randomization: + Verification bias: + Test reliability/variability: + Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: + Testing protocol described: + Sample size: + Statistical tests: + MANAGEMENT Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: - No assessment of cervical ripeness – may explain high rate of meconium and C-section among those women with labor induced for abnormal MPD. Sample size estimates based on differences in cord pH. *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	normal and MBP normal, then f/u visit arranged 3 days later, unless patient had reached 43 wks gestation, in which case labor induced regardless of test results. Labor induced with intravaginal prostaglandins (details NR). Reference standard(s): None Dates: Jan-Dec 1973 Location: Liverpool, UK Setting: University hospital Type(s) of providers: Unspecified OB/GYN, MFM Length of follow-up: None			7) Spontaneous labor: Simple: 41/73 (56%) Complex: 29/72 (40%) p = 0.08 8) Inductions: *Overall:* Simple: 30/73 (41%) Complex: 43/72 (60%) p = 0.04 *For abnormal post-term monitoring:* Simple: 11/73 (15%) Complex: 28/72 (39%) p = 0.002 *43 weeks' gestation:* Simple: 12/73 (16%) Complex: 9/72 (13%) p = 0.66 *Maternal request:* Simple: 4/73 (5%) Complex: 2/72 (3%) p = 0.69 *Other:* Simple: 3/73 (4%) Complex: 4/72 (6%) p = 0.9 9) Normal vaginal delivery: Simple: 58/73 (79%) Complex: 50/72 (69%) p = 0.23 10) Abnormal CTG intrapartum: Simple: 29/73 (40%) Complex: 34/72 (47%) p = 0.36	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Arabin, Snyjders, Mohnhaupt, et al., 1993	Design: Case series, no controls Test(s) studied: Note: Tests 1) *and* 2) applied to all patients in the series (n = 110) 1) Traditional biophysical profile 2) Fetal assessment score consisting of 5 components: FHR pattern; uterine artery resistance by Doppler U/S; carotid artery resistance index by Doppler U/S; fetal tone (movements) by U/S; fetal reflexes (magnitude and speed of movements) by U/S Reference standard(s): Fetal distress (pathological FHR pattern resulting in operative delivery, Apgar score < 7 at 1 minute, or cord blood pH < 7.20) Dates: NR Location: Berlin, Germany Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 110 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 110 Inclusion criteria: Gestational age > 290 days; singleton pregnancy Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry: NR; gestational age > 290 days required for entry into study; mean gestational age at delivery 295 days (range, 293-300) (all patients delivered within ≤ 3 days of assessment) Dating criteria: LMP confirmed by "early" U/S Parity: NR Bishop score: NR	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Cord pH < 7.20 4) C-sections due to fetal distress 5) Test performance	1) Apgar score < 7 at 1 minute: 10/110 (9%) 2) Apgar score < 7 at 5 minutes: 2/110 (2%) 3) Cord pH < 7.20: 9/110 (8%) 4) C-sections due to fetal distress: 38/110 (34.5%) 5) Test performance: Fetal assessment score provided better prediction of fetal distress and low Apgar score at 1 minute than did biophysical profile in ROC analysis ($p < 0.001$). No difference between the two tests for prediction of low pH. Stepwise discriminant analysis of individual components of biophysical profile showed that only FHR pattern and AFV contributed significantly to the diagnostic properties of the total score. Similar analysis of the new fetal assessment score showed that all components except fetal tone contributed significantly to the diagnostic properties of the total score.	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: - Dating criteria: + Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: + Statistical tests: + Fetal assessment score most superior to biophysical profile score in discriminating the relatively subjective outcome of "fetal distress."

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Arias, 1987	Design: RCT, randomization by last digit of year of birth Test(s) studied: 1) Nonstress test (NST) (n = 126) Protocol: Patients evaluated with weekly NST. NST considered reactive if 5 or more accelerations of ≥ 15 bpm lasting at least 15 sec, in association with fetal movements, in 20 minutes. If NST nonreactive, then oxytocin challenge test (OCT) performed. If OCT positive or suspicious, then labor induced. Method of induction not described. 2) U/S + NST (n = 117) Protocol: Weekly U/S evaluation, with assessment of fetal weight, AFV, and placenta. If placenta was grade III and there was decreased AFV, or if fetal weight ≥ 4000 g, then labor induced. Weekly NST as above, with same criteria for induction. Method of induction not described. Reference standard(s): Occurrence of abnormal outcomes (except those not predictable by NST) Dates: NR (15 months' duration) Location: St. Louis, MO Setting: Community hospital	No. of subjects at start: 287 Dropouts: 44 Loss to follow-up: NA No. of subjects at end: 243 Inclusion criteria: Excellent dates (based on LMP or U/S); > 40 wks gestation Exclusion criteria: Diabetes; hypertension; any medical complication of pregnancy Age (mean ± SD): NST: 25.6 ± 4.9; U/S + NST: 25.9 ± 4.9 Race: NR Gestational age at entry (mean ± SD): NST: 41.2 ± 0.7 weeks; U/S + NST: 41.2 ± 0.6 weeks Dating criteria: LMP or U/S during first 26 weeks Parity (mean ± SD): NST: 1.8 ± 1.1; U/S + NST: 1.8 ± 1.2 Bishop score: NR	1) Mean birthweight 2) Birthweight > 4000 g 3) Birthweight > 4500 g 4) Any complication 5) Shoulder dystocia 6) Meconium aspiration 7) Post-maturity syndrome 8) C-sections 9) C-sections due to fetal distress 10) 2 x 2 tables	P-values not reported for the outcomes listed here. 1) Mean birthweight (± SD): NST: 3742 ± 472 g U/S + NST: 3813 ± 482 g 2) Birthweight > 4000 g: NST: 45/126 (36%) U/S + NST: 27/117 (23%) 3) Birthweight > 4500 g: NST: 10/126 (8%) U/S + NST: 9/117 (8%) 4) Any complication: NST: 32/126 (25%) U/S + NST: 29/117 (25%) 5) Shoulder dystocia: NST: 6/126 (5%) U/S + NST: 2/117 (2%) 6) Meconium aspiration: NST: 5/126 (4%) U/S + NST: 3/117 (3%) 7) Post-maturity syndrome: NST: 5/126 (4%) U/S + NST: 4/117 (3%) 8) C-sections: NST: 32/126 (25%) U/S + NST: 33/117 (28%) 9) C-sections due to fetal distress: NST: 12/126 (9.5%) U/S + NST: 16/117 (14%)	QUALITY SCORES: TESTING Reference standard: + Randomized: + Method of randomization: - Verification bias: + Test reliability/variability: + Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: - MANAGEMENT Randomized: + Method of randomization: - Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: -

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Type(s) of providers: Unspecified OB/GYN Length of follow-up: None			10) 2 x 2 tables: 2 x 2 Table 1: Reference standard = abnormal outcomes Screening test = NST Abnormal outcomes yes / no / Totals: NST + 6 / 8 / 14 NST - 12 / 86 / 98 Totals: 18 / 94 / 112 2 x 2 Table 2: Reference standard = abnormal outcomes Screening test = U/S + NST Abnormal outcomes yes / no / Totals: NST + 15 / 15 / 30 NST - 26 / 49 / 75 Totals: 41 / 64 / 105	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Battaglia, Larocca, Lanzani, et al., 1991	Design: Case series (prospective), no controls Test(s) studied: 1) Nonstress test (NST) + amnioscopy + amniotic fluid volume (AFV) + Doppler velocimetry of the uterine, umbilical, descending thoracic aorta, renal, and middle cerebral arteries + hPL + estriol + hematocrit + platelets + mean platelet volume + uric acid Protocol: NST, amnioscopy, AFV, and Doppler velocimetry performed every other day; remaining tests performed every 3 days. Time-averaged mean velocity in the descending thoracic aorta calculated using mean value of three consecutive waveforms. Reference standard(s): 1) "Poor condition" 2) Oligohydramnios 3) Meconium staining 4) NST 5) C-sections (overall) 6) C-sections for fetal distress Dates: Jan - Dec 1989 Location: Modena, Italy Setting: University hospital Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: 82 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 82 Inclusion criteria: Gestational age ≥ 287 days; singleton fetus; cephalic presentation Exclusion criteria: Medical or obstetric complications Age (mean, with range): 27.9 (19-39) Race: NR Gestational age at entry (mean): 292.4 days Dating criteria: LMP + U/S before 24 weeks Parity: 0: 58/82 (71%) 1: 18/82 (22%) > 1: 6/82 (7%) Bishop score: NR	1) Birthweight 2) Macrosomia (birthweight > 4000 g) 3) "Poor condition" (both 1- and 5-minute Apgar scores < 7 *or* infant admitted to NICU for asphyxia and/or meconium aspiration syndrome) 4) Oligohydramnios (largest pocket < 2 cm) 5) Meconium staining 6) C-sections 7) 2 x 2 tables	1) Birthweight (mean): 3655.5 g 2) Macrosomia: 18/82 (22%) 3) "Poor condition": 1/82 (1%) 4) Oligohydramnios: 25/82 (30%) 5) Meconium staining: 24/82 (29%) 6) C-sections: 24/82 (29%) 7) 2 x 2 tables: <u>2 x 2 table 1:</u> Reference standard = "Poor condition" (as defined at left) Screening test = Time-averaged mean velocity of the descending thoracic aorta ("normal" defined as > 25 cm/sec) Poor condition yes / no / Totals: Velocity abnormal: 1 / 23 / 24 Velocity normal: 0 / 58 / 58 Totals: 1 / 81 / 82 Sensitivity: 100% Specificity: 71% <u>2 x 2 table 2:</u> Reference standard = Oligohydramnios Screening test = Time-averaged mean velocity of the descending thoracic aorta ("normal" defined as > 25 cm/sec) Oligohydramnios yes / no / Totals: Velocity abnormal: 16 / 8 / 24 Velocity normal: 9 / 49 / 58 Totals: 25 / 57 / 82	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: - *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				Sensitivity: 64% Specificity: 86% 2 x 2 table 3: Reference standard = Meconium staining Screening test = Time-averaged mean velocity of the descending thoracic aorta ("normal" defined as > 25 cm/sec) Meconium: yes / no / Totals: Velocity abnormal: 22 / 2 / 24 Velocity normal: 2 / 56 / 58 Totals: 24 / 58 / 82 Sensitivity: 92% Specificity: 97% 2 x 2 table 4: Reference standard = NST Screening test = Time-averaged mean velocity of the descending thoracic aorta ("normal" defined as > 25 cm/sec) NST: abn / nl / Totals: Velocity abnormal: 13 / 11 / 24 Velocity normal: 0 / 58 / 58 Totals: 13 / 69 / 82 Sensitivity: 100% Specificity: 84% 2 x 2 table 5: Reference standard = C-sections (overall) Screening test = Time-averaged mean velocity of the descending thoracic aorta ("normal" defined as > 25 cm/sec)	*(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

	C-section yes	C-section no	Totals:
Velocity abnormal	14	10	24
Velocity normal	10	48	58
Totals:	24	58	82

Sensitivity: 58%
Specificity: 50%

2 x 2 table 6:
Reference standard = C-section for fetal distress
Screening test = Time-averaged mean velocity of the descending thoracic aorta ("normal" defined as > 25 cm/sec)

	C-section/fetal distress yes	C-section/fetal distress no	Totals:
Velocity abnormal	8	16	24
Velocity normal	2	56	58
Totals:	10	72	82

Sensitivity: 80%
Specificity: 78%

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Bochner, Medearis, Ross, et al., 1987	Design: Cohort study Test(s) studied: 1) Antepartum testing, including amniotic fluid assessment, NST, and, when necessary, contraction stress testing (CST). Uterine contractions, FHR, and fetal movements also assessed. Protocol: Testing performed twice weekly. Abnormal testing, leading to induction, included decreased amniotic fluid; repetitive variable or late decelerations during the NST or CST; and a nonreactive NST in a patient with an inducible cervix. Patients with a nonreactive NST and an unfavorable cervix had a repeat NST 2 hours later. CST done if the NST was again nonreactive. If the CST negative, then patients re-tested in 3-4 days. Reference standard(s): 1) Meconium aspiration 2) Low birthweight (< 10th percentile) 3) Perinatal mortality or morbidity 4) C-section for fetal distress 5) Apgar < 7 at 1 minute 6) Apgar < 7 at 5 minutes Dates: Jan 1983 - Jan 1986 Location: Los Angeles, CA Setting: University hospital Type(s) of providers:	No. of subjects at start: 845 Dropouts: 6 Loss to follow-up: NA No. of subjects at end: 839 Inclusion criteria: Gestational age of 41-42 completed weeks; referred for post-term fetal assessment Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry: NR (gestational age of 41-42 completed weeks required for entry into study) Dating criteria: Combinations of early dating criteria, including LMP, initial uterine exam, 1st or 2nd trimester U/S, and timing of initial fetal heart tones by Doppler or fetoscopic auscultation Parity: NR Bishop score: NR	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Meconium aspiration 4) Mortality 5) Low birthweight (< 10th percentile) 6) C-section for fetal distress 7) 2 x 2 tables	1) Apgar score < 7 at 1 minute: 56/83 (6.7%) 2) Apgar score < 7 at 5 minutes: 13/839 (1.5%) 3) Meconium aspiration: 3/839 (0.4%) 4) Mortality: 0/839 5) Low birthweight (< 10th percentile): 7/839 (0.8%) 6) C-section for fetal distress: 52/839 (6.2%) 7) 2 x 2 tables (for patients with heavy meconium at rupture of the membranes only [n = 62]): 2 x 2 table 1: Reference standard = Meconium aspiration Screening test = Antepartum testing Meconium aspiration yes / no / Totals: Antepartum testing abn: 1 / 13 / 14 Antepartum testing nl: 2 / 46 / 48 Totals: 3 / 59 / 62 2 x 2 table 2: Reference standard = Low birthweight (defined as < 10th percentile) Screening test = Antepartum testing Low birthweight yes / no / Totals: Antepartum testing abn: 2 / 12 / 14 Antepartum testing nl: 5 / 43 / 48	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: - Dating criteria: + Other risk factors absent: - Similar to likely pt pop: - Testing protocol described: + Sample size: - Statistical tests: - *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Unspecified OB/GYN Length of follow-up: None			Totals: 7 55 62 2 x 2 table 3: Reference standard = Perinatal mortality or morbidity Screening test = Antepartum testing Mortality/ morbidity: yes / no / Totals: Antepartum testing abn: 0 / 14 / 14 Antepartum testing nl: 0 / 48 / 48 Totals: 0 / 62 / 62 2 x 2 table 4: Reference standard = C-section for fetal distress Screening test = Antepartum testing C-section: yes / no / Totals: Antepartum testing abn: 11 / 3 / 14 Antepartum testing nl: 2 / 46 / 48 Totals: 13 / 49 / 62 2 x 2 table 5: Reference standard = Apgar score at 1 minute Screening test = Antepartum testing Apgar at 1 min: < 7 / ≥ 7 / Totals: Antepartum testing abn: 6 / 8 / 14 Antepartum testing nl: 18 / 30 / 48 Totals: 24 / 38 / 62	

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				2 x 2 table 6: Reference standard = Apgar score at 5 minutes Screening test = Antepartum testing Apgar at 5 min < 7 ≥ 7 Totals: Antepartum testing abn 1 13 14 Antepartum testing nl 0 48 48 Totals: 1 61 62	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Bochner, Williams III, Castro, et al., 1988	Design: Case series, concomitant controls Test(s) studied: 1) Antenatal testing beginning at 41 (n = 908) or 42 (n = 352) weeks Protocol: Testing performed twice weekly. Standard fetal monitor recorded uterine contractions, fetal heart rate, and fetal movements. U/S evaluated AFV (< 3 cm abnormal). Nonstress test (NST) also performed. If NST nonreactive *and* AFV normal *and* cervix unfavorable for induction, then NST repeated in 2 hours; if second NST nonreactive, then contraction stress test (CST) performed. If CST negative, then patient re-tested in 3-4 days. Criteria for induction: Decreased AFV (< 3 cm); *or* bradycardia or repetitive variable or late decelerations during NST or CST; *or* nonreactive NST and inducible cervix. Method of induction not described. 2) No antenatal testing (n = 1807 controls). Management protocol not described. Reference standard(s): Intra-partum fetal distress, defined as: a) repetitive late decelerations; b) repetitive moderate or severe variable decelerations with pH < 7.2 or decreased variability; or c)	No. of subjects at start: 1260 subjects, 1807 controls Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 1260 subjects, 1807 controls Inclusion criteria: Uncomplicated post-term pregnancy (> 41 wks); first seen before 20 wks; trial of labor; delivery within 4 days of antepartum testing Exclusion criteria: High risk factors; suspected fetal growth retardation Age: NR Race: NR Gestational age at entry: NR Dating criteria: Accurate LMP; or 1st trimester uterine exam; or 1st or 2nd trimester U/S; or timing of initial auscultated fetal heart tones Parity: NR Bishop score: NR	1) Apgar scores < 7 at 1 minute 2) Apgar scores < 7 at 5 minutes 3) Meconium aspiration 4) Low birthweight 5) Stillbirth 6) Neonatal death 7) Major neonatal morbidity 8) Elective induction 9) C-sections 10) Total adverse outcomes 11) 2 x 2 tables 12) Predictive values	Outcomes 1-11 reported for subjects who delivered between 41 and 42 weeks (n = 512) and for controls, all of whom (n = 1807) delivered between 41 and 42 weeks. 1) Apgar scores < 7 at 1 minute: Testing: 24/512 (4.7%) No testing: 92/1807 (5.1%) p = not significant 2) Apgar scores < 7 at 5 minutes: Testing: 3/512 (0.6%) No testing: 16/1807 (0.9%) p = not significant 3) Meconium aspiration: Testing: 0/512 No testing: 3/1807 (0.2%) p = not significant 4) Low birthweight (< 10th percentile): Testing: 37/512 (7.2%) No testing: 123/1807 (6.8%) p = not significant 5) Stillbirth: Testing: 0/512 No testing: 3/1807 (0.2%) p = not significant 6) Neonatal death: Testing: 0/512 No testing: 0/1807 p = not significant 7) Major neonatal morbidity: Testing: 0/512 No testing: 7/1807 (0.4%) p = not significant 8) Elective induction: Testing: 62/512 (12%) No testing: 282/1807 (16%)	QUALITY SCORES: TESTING Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: - Dating criteria: + Other risk factors absent: + Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: + MANAGEMENT Randomized: - Method of randomization: NA Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: - Dating criteria: + Bishop score: - *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	prolonged bradycardia Dates: Jan 1984 – Jan 1987 Location: Los Angeles, CA Setting: Community hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None			p = not significant 9) C-sections: *Overall:* Testing: 115/512 (22%) No testing: 396/1807 (22%) p = not significant *For fetal distress:* Testing: 14/512 (2.7%) No testing: 60/1807 (3.3%) p = 0.07 *For other indications:* Testing: 101/512 (20%) No testing: 336/1807 (19%) p = not significant 10) Total number of adverse outcomes: Testing: 0/512 No testing: 13/1807 (0.7%) p < 0.05 11) 2 x 2 tables: 2 x 2 Table 1 (n = 908 subjects who started testing at 41 weeks): Reference standard = Intrapartum fetal distress Screening test = Testing Fetal distress yes / no / Totals: Screen test abn: 16 / 119 / 135 Screen test nl: 7 / 766 / 773 Totals: 23 / 885 / 908 2 x 2 Table 2 (n = 352 subjects who started testing at 42 weeks): Reference standard = Intrapartum fetal distress Screening test = Testing *(continued on next page)*	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

	Fetal distress yes	no	Totals:
Screen test abn	17	60	77
Screen test nl	4	271	275
Totals:	21	331	352

12) Predictive values of testing: Positive predictive value significantly higher for testing at 42 weeks than for testing at 41 weeks (21.1% vs. 11.9%, respectively). Negative predictive value significantly lower for testing at 42 weeks than for testing at 41 weeks (98.5% vs. 99.1%, respectively).

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Brar, Horenstein, Medearis, et al., 1989	Design: Case series (prospective), no controls Test(s) studied: 1) Nonstress test (NST) + amniotic fluid volume (AFV) assessment + vascular resistance as measured by Doppler U/S (n = 45) Protocol: NST and AFV performed twice weekly. Reactive NST defined as two accelerations in a 10-minute moving window or an acceleration of 15 beats by 15 seconds. AFV > 5 cm considered normal. Flow velocity waveforms of the left and right uterine artery and the umbilical artery obtained with a continuous wave Doppler U/S. Peak systolic (S) to end-diastolic (D) ratios computed over three different cardiac cycles; mean value calculated and used for analysis. Umbilical artery S:D ratio > 3 considered abnormal, as was any diastolic notching. Uterine artery S:D ratio > 2.6 considered abnormal. Reference standard(s): 1) C-section for fetal distress 2) Meconium 3) Apgar score at 5 minutes 4) Admission to NICU 5) Dysmature Dates: NR Location: Los Angeles, CA Setting: University hospital	No. of subjects at start: 45 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 45 Inclusion criteria: Gestational age ≥ 287 days Exclusion criteria: Medical or obstetric complication Age: NR Race: NR Gestational age at entry: NR (gestational age ≥ 287 days required for entry into study) Dating criteria: LMP confirmed by one of following: early pregnancy test; 1st trimester exam; U/S prior to 24 weeks; or fetal heart tones by fetoscopy at 18-20 weeks Parity: NR Bishop score: NR	1) Apgar score < 7 at 5 minutes 2) Meconium 3) Admission to NICU 4) Dysmature 5) C-section for fetal distress 6) 2 x 2 tables 7) Other test performance results	1) Apgar score < 7 at 5 minutes: 8/45 (18%) 2) Meconium: 11/45 (24%) 3) Admission to NICU: 6/45 (13%) 4) Dysmature: 3/45 (7%) 5) C-section for fetal distress: 13/45 (29%) 6) 2 x 2 tables: 2 x 2 table 1: Reference standard = C-section for fetal distress Screening test = Antepartum testing (APT) (NST and AFV) C-section yes / no / Totals: APT abnormal: 9 / 10 / 19 APT normal: 4 / 22 / 26 Totals: 13 / 32 / 45 2 x 2 table 2: Reference standard = Meconium Screening test = Antepartum testing (APT) (NST and AFV) Meconium yes / no / Totals: APT abnormal: 10 / 9 / 19 APT normal: 1 / 25 / 26 Totals: 11 / 34 / 45	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: - Dating criteria: - Other risk factors absent: - Similar to likely pt pop: - Testing protocol described: - Sample size: - Statistical tests: - Relationship between Doppler studies and fetal outcomes not reported.

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Type(s) of providers: Not specified Length of follow-up: None			2 x 2 table 3: Reference standard = Apgar score at 5 minutes Screening test = Antepartum testing (APT) (NST and AFV)	

	Apgar at 5 min		
	< 7	≥ 7	Totals:
APT abnormal	7	12	19
APT normal	1	25	26
Totals:	8	37	45

2 x 2 table 4:
Reference standard = Admission to NICU
Screening test = Antepartum testing (APT) (NST and AFV)

	NICU admission		
	yes	no	Totals:
APT abnormal	5	14	19
APT normal	1	25	26
Totals:	6	39	45

2 x 2 table 5:
Reference standard = Dysmature
Screening test = Antepartum testing (APT) (NST and AFV)

	Dysmature		
	yes	no	Totals:
APT abnormal	2	17	19
APT normal	1	25	26
Totals:	3	42	45

7) Other test performance results:
Umbilical and uterine artery S:D ratios were not significantly different between patients with normal and abnormal

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				antepartum test results. Cerebral S:D and cerebral placental resistance ratios were significantly lower in patients with abnormal antepartum test results.	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Chauhan, Sullivan, Lutton, et al., 1995	Design: Case series, no controls Test(s) studied: 1) Maternal estimation of birthweight (n = 70) Protocol: Patients interviewed as follows: "With your previous deliveries you looked and felt a certain way, and the newborn(s) weighed X amount. Based solely on those experiences, how much do you think this newborn will weigh?" 2) Clinical estimation of birthweight (n = 40) Protocol: Performed by obstetrician or midwife using Leopold's maneuvers alone (no computations or formulas). Reference standard(s): Actual birthweight Dates: NR; study conducted over a 3-year period Location: NR Setting: 3 unspecified hospitals Type(s) of providers: Unspecified OB/GYNs (n = 3); unspecified midwives (n = 2) Length of follow-up: None	No. of subjects at start: 70, all of whom provided maternal estimation of birthweight, and 40 of whom also received clinical estimation of birthweight Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 70 Inclusion criteria: Gestational age ≥ 41 weeks; parous; in early active labor with singleton gestation; vertex presentation; no evidence of fetal distress Exclusion criteria: None specified Age (mean ± SD, with range): 26.1 ± 4.5 (range, 17-38) Race: NR Gestational age at entry (mean ± SD, with range): 41.5 ± 0.6 weeks (range, 41-43 weeks) Dating criteria: LMP plus early obstetric examination *or* U/S before 20 weeks Parity (mean ± SD, with range): 1.4 ± 0.6 (range, 1-4) Bishop score: NR	1) Absolute error of birthweight estimate (absolute value of estimate - actual birthweight) 2) Standardized error of birthweight estimate (absolute error [g]/actual birthweight [kg]) 3) Percentage of estimates within ± 10% of actual birthweight 4) Sensitivity, specificity, and positive and negative predictive values of estimates ≥ 4000 g for predicting actual birthweight ≥ 4000g 5) Incidence of macrosomia (birthweight ≥ 4000 g)	1) Absolute error of birthweight estimate (mean ± SD; n = 40 women with both maternal and clinical estimates): Clinical estimate: 278 ± 232 g Maternal estimate: 349 ± 331 g p = not significant 2) Standardized error of birthweight estimate (mean ± SD; n = 40 women with both maternal and clinical estimates): Clinical estimate: 75 ± 71 g Maternal estimate: 92 ± 81 g p = not significant 3) Percentage of estimates within ± 10 of actual birthweight (mean ± SD; n = 40 women with both maternal and clinical estimates): Clinical estimate: 65.0% Maternal estimate: 67.5% p = not significant 4) Sensitivity, specificity, and positive and negative predictive values of estimates ≥ 4000 g for predicting actual birthweight ≥ 4000g: Maternal estimates (n = 70): Sensitivity: 56% Specificity: 94% + predictive value: 77% - predictive value: 86% Clinical estimates (n = 40): Sensitivity: 62% Specificity: 92% + predictive value: 70% - predictive value: 82% 5) Incidence of macrosomia: 18/70 (25.7%)	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: + Dating criteria: + Other risk factors absent: - Similar to likely pt pop: ? Testing protocol described: + Sample size: - Statistical tests: + Differential sample size – 70 for maternal estimates vs. 40 for clinical estimates.

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Chauhan, Sullivan, Magann, et al., 1994	Design: Case series (prospective), no controls Test(s) studied: Note: Birthweight estimated for each participant using both of the following methods: 1) Clinical estimate of birthweight Protocol: Estimated in early labor by clinician using Leopold maneuvers. 2) Sonographic estimate of birthweight Protocol: Same clinician obtained standard sonographic measurements of transverse abdominal diameter, anteroposterior abdominal diameter, and femur length, also in early labor. Reference standard(s): 1) Actual birthweight Dates: NR; study conducted over a 2-year period Location: Jackson, MS Setting: Community hospital Type(s) of providers: MFM Length of follow-up: None	No. of subjects at start: 84 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 84 Inclusion criteria: Gestational age ≥ 41 weeks Exclusion criteria: None specified Age (mean ± SD): 25.9 ± 4.7 Race: NR Gestational age at entry: NR (gestational age ≥ 41 weeks required for entry into study) Dating criteria: LMP + physical exam in 1st trimester or U/S at 20 weeks or earlier Parity (mean ± SD): 0.6 ± 0.7 Bishop score: NR	1) Mean absolute error of the two methods of estimating birthweight 2) Mean percentage absolute error 3) Percentage of estimates within 10% of actual birthweight 4) 2 x 2 tables	1) Mean absolute error of the two methods of estimating birthweight (± SD): Clinical: 322 ± 253 g Sonographic: 547 ± 425 g p < 0.001 2) Mean percentage absolute error (± SD): Clinical: 8.9 ± 7.1 g/kg Sonographic: 14.8 ± 11.0 g/kg p < 0.001 3) Percentage of estimates within 10% of actual birthweight: Clinical: 65.4% Sonographic: 42.8% p < 0.005 4) 2 x 2 tables: 2 x 2 table 1: Reference standard = Actual birthweight Screening test = Clinical estimate of birthweight Actual birthweight ≥ 4000 g / < 4000 g / Totals: Clin est ≥ 4000 g: 10 / 2 / 12 Clin est < 4000 g: 10 / 62 / 72 Totals: 20 / 64 / 84 2 x 2 table 2: Reference standard = Actual birthweight Screening test = Sonographic estimate of birthweight Actual birthweight ≥ 4000 g / < 4000 g / Totals: Sonog est ≥ 4000 g: 11 / 6 / 17 Sonog est < 4000 g: 9 / 58 / 67 Totals: 20 / 64 / 84	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: + Gestational age: + Dating criteria: + Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: +

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Chervenak, Divon, Hirsch, et al., 1989	Design: Case series (not specified if prospective or retrospective), with concomitant controls Test(s) studied: 1) Nonstress test (NST) + amniotic fluid volume (AFV) assessment + U/S estimation of fetal weight (n = 317 cases) Protocol: NST and AFV performed twice weekly. Fetal weight estimated (timing not specified) by biparietal diameter, femur length, and abdominal circumference. Estimated weight did not determine management. Reference standard(s): 1) Actual birthweight Dates: Jan 1987- June 1988 Location: NR Setting: Community hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up:	No. of subjects at start: 317 cases; 100 controls (consecutive patients between 38 and 40 weeks gestational age with no antepartum complications) Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 317 cases; 100 controls Inclusion criteria: Singleton, uncomplicated pregnancy; intact membranes; gestational age > 41 weeks Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry (mean ± SD): Cases, 42 ± 0.6 weeks; controls, 39.8 ± 0.5 weeks Dating criteria: LMP plus early first examination and U/S at < 20 weeks Parity: NR Bishop score: NR	1) Birthweight (mean) 2) Birthweight > 4000 g 3) C-sections 4) 2 x 2 table 5) Other test performance results	1) Birthweight (mean ± SD): Among study patients who delivered at 41 completed weeks (n = 172): 3710 ± 452 g Among study patients who delivered at ≥ 42 completed weeks (n = 145): 3705 ± 454 g Among control patients (n = 100): 3339 ± 360 g No p-values reported 2) Birthweight > 4000 g: Study patients: 81/317 (25.6%) Controls: 6/100 (6%) p < 0.05 3) C-sections: *Overall:* Study patients: 76/317 (24.0%) Controls: 4/100 (4%) p < 0.05 *Primary and repeat C-sections (study patients only):* Primary C-sections: 72/317 (22.7%) Repeat C-sections: 4/317 (1.3%) *C-sections for arrest or protraction disorders (study patients only):* Birthweights > 4000 g: 18/81 (22%) Birthweights < 4000 g: 23/235 (10%) p < 0.01 4) 2 x 2 table: Reference standard = Actual birthweight Screening test = Estimated birthweight (EBW) Actual birthweight |	> 4000 g	< 4000 g	Totals: | | EBW > 4000 g | 49 | 22 | 71 | | EBW < 4000 g | 32 | 214 | 246 | | Totals: | 81 | 236 | 317 |	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: + Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: - Unclear whether estimated fetal weight available to practitioner – possibility of bias in outcome of C-section. Morbidity related to macrosomia not reported. *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				5) Other test performance results: *Performance characteristics of estimated birthweight > 4000 g for predicting actual birthweight > 4000 g:* Sensitivity, 61%; specificity, 91%; positive predictive value, 70%; negative predictive value, 87% *Percentage of estimates within 15% of actual birthweight:* When based on biparietal diameter and abdominal circumference: 88% When based on biparietal diameter and femur length: 87% *Percentage of estimates within 10% of actual birthweight:* When based on biparietal diameter and abdominal circumference: 70% When based on biparietal diameter and femur length: 68% *Mean percentage error of estimates (± SD):* 7.5% ± 6.4%	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Crowley, O'Herlihy, and Boylan, 1984	Design: Cohort study Test(s) studied: 1) U/S assessment of AFV Protocol: AFV assessed at 42 weeks and every 4 days thereafter until delivery. If AFV reduced (no vertical pool measuring > 3 cm), then labor induced by amniotomy and oxytocin 24 hours later, if needed. Reference standard(s): 1) Meconium staining 2) C-section for fetal distress 3) Low birthweight (< 10th percentile) 4) Admission to NICU Dates: NR Location: Dublin, Ireland Setting: Unspecified hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 335 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 335 Inclusion criteria: Singleton pregnancy at 42 weeks Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry: 42 weeks Dating criteria: Certain LMP or early U/S Parity: 138/335 (41%) primigravidae; 197/335 (59%) multigravidae Bishop score: NR	1) Meconium, grade I 2) Meconium, grade II or III 3) Low birthweight (< 10th percentile) 4) Admission to NICU 5) Convulsions 6) Abnormal tone and primitive reflexes 7) C-sections 8) 2 x 2 tables 9) Other test performance results	1) Meconium, grade I: 24/335 (7%) 2) Meconium, grade II or III: 24/335 (7%) 3) Low birthweight (< 10th percentile): 37/335 (11%) 4) Admission to NICU: 24/335 (7%) 5) Convulsions: 0/335 6) Abnormal tone and primitive reflexes: 2/335 (< 1%) 7) C-sections: 26/335 (8%) Overall: 26/335 (8%) For fetal distress: 9/335 (3%) For dystocia: 8/335 (2%) For failed induction: 3/335 (< 1%) Elective: 6/335 (2%) 8) 2 x 2 tables: <u>2 x 2 Table 1:</u> Reference standard = C-section for fetal distress Screening test = AFV (abn < 3 cm; nl > 3 cm) |	C-section	No C-section	Totals:
---	---	---	---
AFV abn	7	58	65
AFV nl	2	268	270
Totals:	9	326	335	 <u>2 x 2 Table 2:</u> Reference standard = Low birthweight (BW) (< 10th percentile) Screening test = AFV (abn < 3 cm; nl > 3 cm)	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: + Test reliability/variability: + Gestational age: + Dating criteria: + Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: +

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

BW

	BW low	BW not low	Totals:
AFV abn	17	48	65
AFV nl	20	250	270
Totals:	37	298	335

2 x 2 Table 3:
Reference standard = Admission to NICU
Screening test = AFV (abn < 3 cm; nl > 3 cm)

	NICU yes	NICU no	Totals:
AFV abn	9	56	65
AFV nl	15	255	270
Totals:	24	311	335

9) Other test performance results:
Clinical assessment of AFV by abdominal palpation showed a false positive rate of 25% and a false negative rate of 43% for detecting "significant meconium staining or absent amniotic fluid." Sensitivity, 75%; specificity, 57%.

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Devoe and Sholl, 1983	Design: Case series, no controls Test(s) studied: 1) Maternal estriol + fetal heart rate tests (NST and CST) Protocol: Serial maternal urinary or plasma estriol tests performed biweekly. NST performed weekly and considered reactive if 3 or more accelerations of > 15 bpm amplitude and 15-second duration occurred, with fetal movements, in 30 minutes. If NST nonreactive, then CST performed. CST considered positive if at least 30% of contractions, occurring at a rate of 3/10 min, were followed by late decelerations in a 30-min period. CST equivocal if fewer late decelerations occurred and negative if no late decelerations occurred. Labor induced "either for elective reasons or because of abnormal fetal test results." Method of induction not described. Reference standard(s): 1) Apgar score at 5 minutes 2) Intrapartum fetal distress Dates: July 1977-June 1981 Location: Chicago, IL Setting: University hospital	No. of subjects at start: 248 Dropouts: NR Loss to follow-up: NA No. of subjects at end: 248 (if no dropouts) Inclusion criteria: Singleton pregnancy; unripe cervix at 40 weeks Exclusion criteria: Significant medical or OB complications Age: NR Race: NR Gestational age at entry: 40 weeks Dating criteria: Known LMP confirmed by OB milestones, early clinical exam, or U/S Parity: NR Bishop score: NR	1) Meconium staining 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Perinatal mortality 5) Intrauterine growth retardation (IUGR) 6) Post-maturity syndrome 7) Intrapartum fetal distress – defined as presence of 2 or more of the following: (a) persistent fetal tachycardia or bradycardia; (b) loss of beat-to-beat variability; (c) severe variable or late decelerations; (d) passage of thick, fresh meconium; or (e) scalp pH < 7.22 8) C-sections 9) 2 x 2 tables	1) Meconium staining: 74/248 (30%) 2) Apgar score < 7 at 5 minutes: 7/248 (3%) 3) Birthweight (mean ± SD): 3418 ± 443 g 4) Perinatal mortality: 2/248 (<1%) 5) IUGR: 7/248 (3%) 6) Post-maturity syndrome: 13/248 (5%) 7) Intrapartum fetal distress: 43/248 (17%) 8) C-sections: 34/248 (14%) 9) 2 x 2 tables: 2 x 2 Table 1: Reference standard = Apgar score at 5 minutes Screening test = FHR tests (NST and CST) Apgar < 7 / Apgar ≥ 7 / Totals: NST non-r CST pos: 0 / 22 / 22 NST non-r CST neg: 0 / 17 / 17 NST r (no CST): 7 / 202 / 209 Totals: 7 / 241 / 248	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: -

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Type(s) of providers: Unspecified OB/GYN Length of follow-up: None			2 x 2 Table 2: Reference standard = Fetal distress (yes/no) Screening test = FHR tests (NST and CST) Distress yes / Distress no / Totals: NST non-r CST pos: 6 / 16 / 22 NST non-r CST neg: 6 / 11 / 17 NST r (no CST): 31 / 178 / 209 Totals: 43 / 205 / 248 2 x 2 Table 3: Reference standard = Apgar score at 5 minutes Screening test = Maternal estriol ("low" = below the 10th percentile for gestational age; "falling" = drop of more than 40% from mean of the 3 highest preceding values) Apgar < 7 / Apgar ≥ 7 / Totals: Estriol low or falling: 0 / 46 / 46 Estriol nl: 6 / 166 / 172 Totals: 6 / 212 / 218 2 x 2 Table 4: Reference standard = Fetal distress (yes/no) Screening test = Maternal estriol (as above) Distress yes / Distress no / Totals: Estriol low or falling: 4 / 42 / 46 Estriol nl: 31 / 141 / 172 Totals: 35 / 183 / 218	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Eden, Gergely, Schifrin, et al., 1982	Design: Case series (prospective), no controls Test(s) studied: 1) NST + CST (n = 78) Protocol: Weekly NST. If NST nonreactive, then CST. If CST negative, then repeat NST in 1 week. If CST suspicious, then repeat NST in 1 day. If CST positive, then deliver. 2) NST + modified biophysical profile (MBP) (n = 398) Protocol: Semi-weekly NST. If NST nonreactive, then MBP performed. If MBP normal, then NST repeated semi-weekly. If MBP abnormal, then deliver. 3) NST + AFV + MBP (n = 109) Protocol: Semi-weekly NST + weekly AFV. If AFV decreased, then deliver. If NST nonreactive and AFV normal, then perform MBP. If MBP normal, then resume semi-weekly NST and weekly AFV. If MBP abnormal, then deliver. Reference standard(s): 1) Apgar scores at 1 minute 2) Apgar scores at 5 minutes 3) Meconium aspiration 4) Resuscitation 5) C-section Dates: Nov 1978 – Aug 1981 Location: Los Angeles, CA	No. of subjects at start: 585 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 585 Inclusion criteria: 42 weeks of gestation; prenatal care for ≥ 20 weeks Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry: NR Dating criteria: LMP with consistent exams, *or* sequential U/S exams Parity: NR Bishop score: NR	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Meconium aspiration 4) Resuscitation 5) Fetal distress requiring intervention (persistent abnormal FHR patterns) 6) Morbidity (defined as presence of any of following: fetal distress requiring intervention, 5-minute Apgar score < 7, neonatal resuscitation, postmaturity syndrome, meconium aspiration) 7) C-sections 8) 2 x 2 tables	1) Apgar score < 7 at 1 minute: NST + CST: 15.4% NST + MBP: 13.1% NST + AFV + MBP: 7.3% no significant differences 2) Apgar score < 7 at 5 minutes: NST + CST: 10.3% NST + MBP: 2.3% NST + AFV + MBP: 0 1 vs. 2, p < 0.05 1 vs. 3, p < 0.05 3) Meconium aspiration: NST + CST: 6.4% NST + MBP: 1.3% NST + AFV + MBP: 0 1 vs. 2, p < 0.05 4) Resuscitation: NST + CST: 12.8% NST + MBP: 10.1% NST + AFV + MBP: 0 1 vs. 2, p < 0.05 2 vs. 3, p < 0.05 5) Fetal distress: NST + CST: 21.8% NST + MBP: 4.5% NST + AFV + MBP: 5.5% 1 vs. 2, p < 0.05 1 vs. 3, p < 0.05 6) Morbidity: NST + CST: 25.6% NST + MBP: 14.3% NST + AFV + MBP: 5.5% 1 vs. 2, p < 0.05 1 vs. 3, p < 0.05 2 vs. 3, p < 0.05 7) C-sections: NST + CST: 11.5% NST + MBP: 29.9%	QUALITY SCORES: TESTING Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: - Dating criteria: + Other risk factors absent: - Similar to likely pt pop: - Testing protocol described: + Sample size: - Statistical tests: - MANAGEMENT Randomized: - Method of randomization: NA Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: - Statistical tests: - Gestational age: - Dating criteria: + Bishop score: - Women with complications of pregnancy (e.g., preeclampsia, diabetes, previous stillbirth) NOT excluded. *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Setting: University hospital Type(s) of providers: General OB/GYN Length of follow-up: None			NST + AFV + MBP: 29.4% 1 vs. 2, $p < 0.05$ 1 vs. 3, $p < 0.05$ 8) 2 x 2 tables: 2 x 2 Table 1: Reference standard = Apgar score at 1 minute Screening test = AFV For patients in NST + CST group only (n = 78) Apgar at 1 min AFV: < 7 / ≥ 7 / Totals: decreased: 7 / 20 / 27 AFV nl: 6 / 45 / 51 Totals: 13 / 65 / 78 2 x 2 Table 2: Reference standard = Apgar score at 1 minute Screening test = AFV For patients in NST + MBP group only (n = 109) Apgar at 1 min AFV: < 7 / ≥ 7 / Totals: decreased: 4 / 22 / 26 AFV nl: 4 / 79 / 83 Totals: 8 / 101 / 109 2 x 2 Table 3: Reference standard = Apgar score at 5 minutes Screening test = AFV For patients in NST + CST group only (n = 78) Apgar at 5 min AFV: < 7 / ≥ 7 / Totals: decreased: 7 / 20 / 27 AFV nl: 1 / 50 / 51 Totals: 8 / 70 / 78	

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

2 x 2 Table 4:
Reference standard = Apgar score at 5 minutes
Screening test = AFV
For patients in NST + MBP group only (n = 109)

	Apgar at 5 min < 7	Apgar at 5 min ≥ 7	Totals:
AFV decreased	0	26	26
AFV nl	0	83	83
Totals:	0	109	109

2 x 2 Table 5:
Reference standard = Meconium aspiration
Screening test = AFV
For patients in NST + CST group only (n = 78)

	Meconium aspiration yes	Meconium aspiration no	Totals:
AFV decreased	4	23	27
AFV nl	1	50	51
Totals:	5	73	78

2 x 2 Table 6:
Reference standard = Meconium aspiration
Screening test = AFV
For patients in NST + MBP group only (n = 109)

	Meconium aspiration yes	Meconium aspiration no	Totals:
AFV decreased	0	26	26
AFV nl	0	83	83
Totals:	0	109	109

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				2 x 2 Table 7: Reference standard = Resuscitation Screening test = AFV For patients in NST + CST group only (n = 78) Resuscitation yes / no / Totals: AFV decreased: 6 / 21 / 27 AFV nl: 2 / 49 / 51 Totals: 8 / 70 / 78 2 x 2 Table 8: Reference standard = C-section Screening test = AFV For patients in NST + CST group only (n = 78) C-section yes / no / Totals: AFV decreased: 9 / 18 / 27 AFV nl: 10 / 41 / 51 Totals: 19 / 59 / 78 2 x 2 Table 9: Reference standard = Apgar score at 1 minute Screening test = FHR decelerations For patients in NST + CST group only (n = 78) Apgar at 1 min < 7 / ≥ 7 / Totals: FHR dec present: 5 / 5 / 10 FHR dec absent: 7 / 61 / 68 Totals: 12 / 66 / 78 2 x 2 Table 10: Reference standard = Apgar score at 5 minutes Screening test = FHR decelerations For patients in NST + CST group only (n = 78)	*(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

Apgar at 5 min

	< 7	≥ 7	Totals:
FHR dec present	4	6	10
FHR dec absent	4	64	68
Totals:	8	70	78

2 x 2 Table 11:
Reference standard = Meconium aspiration
Screening test = FHR decelerations
For patients in NST + CST group only (n = 78)

	Meconium aspiration yes	no	Totals:
FHR dec present	1	9	10
FHR dec absent	5	63	68
Totals:	6	72	78

2 x 2 Table 12:
Reference standard = Resuscitation
Screening test = FHR decelerations
For patients in NST + CST group only (n = 78)

	Resuscitation yes	no	Totals:
FHR dec present	5	5	10
FHR dec absent	6	62	68
Totals:	11	67	78

2 x 2 Table 13:
Reference standard = C-section
Screening test = FHR decelerations
For patients in NST + CST group only (n = 78)

(continued on next page

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

	C-section yes	no	Totals:
FHR dec present	2	8	10
FHR dec absent	7	61	68
Totals:	9	69	78

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Farmakides, Schulman, Winter, et al., 1988	Design: Case series, no controls Test(s) studied: 1) Nonstress testing (NST) plus Doppler velocimetry Protocol: Testing interval not specified. Management based on NST, but not Doppler velocimetry. Precise management protocols not described. Reference standard(s): 1) C-section for fetal distress 2) Admission to NICU 3) Small for gestational age Dates: "During 1985" Location: Stony Brook, NY Setting: University hospital Type(s) of providers: MFM Length of follow-up: None	No. of subjects at start: 140 (46 of whom were "post-dates") Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 140 Inclusion criteria: Women referred for pre-natal testing for a variety of indications Exclusion criteria: None Age: NR Race: NR Gestational age at entry: NR Dating criteria: NR Parity: NR Bishop score: NR Other: Indications for prenatal testing: Post-dates: 46 (33%) Hypertension: 33 (24%) Diabetes: 14 (10%) Suspected IUGR: 10 (7%) Congenital anomaly: 4 (3%) Other: 33 (24%)	1) Fetal distress (not defined) 2) Small for gestational age (not defined) 3) Admission to NICU 4) C-section for fetal distress 5) 2 x 2 tables	1) Fetal distress: 41/140 (29%). "Most" of the cases of fetal distress came from the post-dates subgoup. 2) Small for gestational age: 15/140 (11%) 3) Admission to NICU: 24/140 (17%) 4) C-section for fetal distress: 39/140 (28%). In the group with abnormal NST, but normal velocimetry, there were significantly more women undergoing C-sections for fetal distress. Again, the majority of these women were in the post-dates subgroup. 5) 2 x 2 tables: 2 x 2 table 1: Reference standard = C-section for fetal distress Screening test = Nonstress test (NST) C-section yes / no / Totals: NST abn 26 / 34 / 60 NST nl 13 / 67 / 80 Totals: 39 / 101 / 140 2 x 2 table 2: Reference standard = NICU admission Screening test = Nonstress test (NST) NICU admission yes / no / Totals: NST abn 13 / 47 / 60 NST nl 11 / 69 / 80 Totals: 24 / 116 / 140 2 x 2 table 3: Reference standard = Small for gestational age (SGA) (not defined) Screening test = Nonstress test (NST) *(continued on next page)*	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: - Dating criteria: - Other risk factors absent: - Similar to likely pt pop: - Testing protocol described: - Sample size: - Statistical tests: - Results not reported separately for subgroup of patients referred for pre-natal testing for "post-date" pregnancy (33% of total study population).

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

	SGA	Not SGA	Totals:
NST abn	9	51	60
NST nl	6	74	80
Totals:	15	125	140

2 x 2 table 4:
Reference standard = NST
Screening test = Velocimetry

	NST abn	NST nl	Totals:
Velocimetry abnormal	16	28	44
Velocimetry normal	44	52	96
Totals:	60	80	140

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Fleischer, Schulman, Farmakides, et al., 1985	Design: Cohort study (retrospective) Test(s) studied: 1) Nonstress testing (NST) (n = 228) Protocol: NST started at 41 weeks' gestation. If score normal (7-10), then NST repeated weekly. If score inconclusive (5-6), then test repeated within 24 hours or followed by a contraction stress test. Patients with abnormal scores on NST (1-4) were evaluated for delivery. 2) No monitoring (n = 30) Protocol: No antenatal monitoring or NST within 7 days of delivery. Reference standard(s): 1) C-section for fetal distress 2) Apgar ≤ 6 at 1 minute 3) Apgar ≤ 6 at 5 minutes 4) Admission to NICU 5) Neonatal death 6) Stillbirth Dates: Jan 1980 - June 1981 Location: Bronx, NY Setting: University hospital Type(s) of providers: General OB/GYN; MFM Length of follow-up: None	No. of subjects at start: 258 (228 of whom received NST within 7 days of delivery) Dropouts: NA (retrospective study) Loss to follow-up: NA No. of subjects at end: 258 Inclusion criteria: Gestational age ≥ 42 weeks at time of delivery Exclusion criteria: Twin gestation; breech presentation; congenital anomalies; chorioamnionitis Age: NR Race: NR Gestational age at entry: NR (NST initiated at 41 weeks) Dating criteria: Consistency between uterine size and gestational age by LMP Parity: NR Bishop score: NR	1) Apgar score ≤ 6 at 1 minute 2) Apgar score ≤ 6 at 5 minutes 3) Admission to NICU 4) Neonatal death 5) Stillbirth 6) C-section for fetal distress 7) 2 x 2 tables	1) Apgar score ≤ 6 at 1 minute: NST: 15/228 (7%) No monitoring: 15/30 (50%) p < 0.01 2) Apgar score ≤ 6 at 5 minutes: NST: 7/228 (3%) No monitoring: 6/30 (20%) p < 0.05 3) Admission to NICU: NST: 7/228 (3%) No monitoring: 5/30 (17%) p < 0.05 4) Neonatal death: NST: 3/228 (1%) No monitoring: 1/30 (3%) p = not significant 5) Stillbirth: NST: 2/228 (1%) No monitoring: 4/30 (13%) p < 0.05 6) C-section for fetal distress: NST: 26/228 (11%) No monitoring: 19/30 (63%) p < 0.001 7) 2 x 2 tables (for patients in the NST group only, n = 228) 2 x 2 table 1: Reference standard = C-section for fetal distress Screening test = Nonstress test (NST) C-section yes / no / Totals: NST 1-4: 6 / 4 / 10 NST 5-6: 5 / 23 / 28 NST 7-10: 15 / 175 / 190 Totals: 26 / 202 / 228	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: + Test reliability/variability: + Gestational age: + Dating criteria: + Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: + *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

2 x 2 table 2:
Reference standard = Apgar score at 1 minute
Screening test = Nonstress test (NST)

	Apgar at 1 min ≤ 6	> 6	Totals:
NST 1-4	4	6	10
NST 5-6	2	26	28
NST 7-10	9	181	190
Totals:	15	213	228

2 x 2 table 3:
Reference standard = Apgar score at 5 minutes
Screening test = Nonstress test (NST)

	Apgar at 5 min ≤ 6	> 6	Totals:
NST 1-4	1	9	10
NST 5-6	1	27	28
NST 7-10	5	185	190
Totals:	7	221	228

2 x 2 table 4:
Reference standard = Admission to NICU
Screening test = Nonstress test (NST)

	NICU admission yes	no	Totals:
NST 1-4	4	6	10
NST 5-6	0	28	28
NST 7-10	3	187	190
Totals:	7	221	228

2 x 2 table 5:
Reference standard = Neonatal death
Screening test = Nonstress test (NST)

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

Neonatal death

	yes	no	Totals:
NST 1-4	3	7	10
NST 5-6	0	28	28
NST 7-10	0	190	190
Totals:	3	225	228

2 x 2 table 6:
Reference standard = Stillbirth
Screening test = Nonstress test (NST)

Stillbirth

	yes	no	Totals:
NST 1-4	0	10	10
NST 5-6	0	28	28
NST 7-10	2	188	190
Totals:	2	226	228

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Gilby, Williams, and Spellacy, 2000	Design: Cohort study Test(s) studied: U/S within 7 days of delivery to measure abdominal circumference Reference standard(s): Macrosomia (defined using two different thresholds, 4000 g and 4500 g) Dates: 1992-1997 Location: Tampa, FL Setting: Community hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 1996 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 1996 Inclusion criteria: Singleton pregnancies with U/S within 7 days of delivery Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry: NR Dating criteria: NR Parity: NR Bishop score: NR	1) 2 x 2 tables 2) Other test performance results	1) 2 x 2 tables: Article includes ROC curves for performance of different abdominal circumference cutoff points (35 cm and 38 cm) for predicting macrosomia at two thresholds, 4000 g and 4500 g. 2 x 2 tables could be constructed only for the 4500 g macrosomia cutoff point. 2 x 2 Table 1: Reference standard = Macrosomia (birthweight ≥ 4500 g) Screening test = Abdominal circumference (Abd C), cutoff point at 35 cm |	BW ≥ 4500 g	BW < 4500 g	Totals:
---	---	---	---
Abd C ≥ 35 cm	68	683	751
Abd C < 35 cm	1	1244	1245
Totals:	69	1927	1996 | 2 x 2 Table 2: Reference standard = Macrosomia (birthweight ≥ 4500 g) Screening test = Abd C, cutoff point at 38 cm |	BW ≥ 4500 g	BW < 4500 g	Totals:
---	---	---	---
Abd C ≥ 38 cm	37	62	99
Abd C < 38 cm	32	1865	1897
Totals:	69	1927	1996	 2) Other test performance results: Abdominal circumference ≥ 35 cm had the following test performance characteristics: Sensitivity, 98.5%; specificity, 64.5%; negative predictive value, 64.5%	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: - Dating criteria: - Other risk factors absent: - Similar to likely pt pop: - Testing protocol described: + Sample size: - Statistical tests: -

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Gilson, O'Brien, Vera, et al., 1988	Design: Case series (not specified if prospective or retrospective), no controls Test(s) studied: 1) Nonstress test (NST) + biophysical profile (BP) (n = 128) Protocol: Testing started when patient "almost" 42 weeks. NST performed twice weekly, BP weekly at first and twice weekly after 43 weeks. Cervix examined at each visit. If BP score 8-10, then patient given another NST in 3-4 days and a repeat BP in 7 days. If BP score 5-7, then BP repeated in 24 hours; if still abnormal, then patient transferred to hospital for induction. If oligohydramnios, spontaneous decelerations on NST, or score < 4, then patient induced. Patients with BP scores of 8-10 allowed to deliver in birthing center if NST reactive and no indication of fetal distress or failure to progress. Otherwise transferred to hospital for labor and delivery. Reference standard(s): 1) Apgar scores at 1 and 5 minutes 2) Post-maturity syndrome 3) Fetal distress 4) C-section for fetal distress Dates: Jan 1984 - Feb 1986 Location: Brownsville, TX Setting: Freestanding birthing center	No. of subjects at start: 178 Dropouts: 50 (delivered before biophysical profile score assessed) Loss to follow-up: NA No. of subjects at end: 128 Inclusion criteria: Gestational age 42 completed weeks; otherwise low risk; biophysical profile score recorded within 1 week of delivery Exclusion criteria: None specified Age: NR Race: 100% Hispanic Gestational age at entry: NR (gestational age of 42 completed weeks required for entry into study) Dating criteria: Clinical sizing (LMP supported by appropriate fundal heights), stethoscope fetal heart tones (for more than 22 weeks), or 2nd trimester U/S Parity: NR Bishop score: NR	1) 2 x 2 tables	1) 2 x 2 tables: 2 x 2 Table 1: Reference standard = Apgar score at 1 minute Screening test = Biophysical profile score (BPS) Apgar < 7 / Apgar ≥ 7 / Totals: BPS < 8: 2 / 24 / 26 BPS 8-10: 12 / 90 / 102 Totals: 14 / 114 / 128 2 x 2 Table 2: Reference standard = Apgar score at 5 minutes Screening test = BPS Apgar < 7 / Apgar ≥ 7 / Totals: BPS < 8: 0 / 26 / 26 BPS 8-10: 0 / 102 / 102 Totals: 0 / 128 / 128 2 x 2 Table 3: Reference standard = Post-maturity syndrome Screening test = BPS Post-maturity: yes / no / Totals: BPS < 8: 7 / 19 / 26 BPS 8-10: 6 / 96 / 102 Totals: 13 / 115 / 128 2 x 2 Table 4: Reference standard = Fetal distress Screening test = BPS Fetal distress: yes / no / Totals: BPS < 8: 4 / 22 / 26 BPS 8-10: 11 / 91 / 102 Totals: 15 / 113 / 128	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: ? Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: - Study underpowered to detect differences in categorical variables. *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Type(s) of providers: Unspecified OB/GYN; nurse midwives Length of follow-up: None			2 x 2 Table 5: Reference standard = C-section for fetal distress Screening test = BPS C-section for fetal distress yes / no / Totals: BPS < 8: 2 / 24 / 26 BPS 8-10: 1 / 101 / 102 Totals: 3 / 125 / 128	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Hann, McArdle, and Sachs, 1987	Design: Case series, no controls Test(s) studied: Biophysical Profile Score (BPS). Included 6 components: 1) NST; 2) fetal breathing movements; 3) fetal movements; 4) fetal tone; 5) amniotic fluid volume (AFV); and 6) placental grading. Score of 0-2 given to each variable. Abnormal score defined as < 6. Patients with scores of 4-6 managed "on an individualized basis"; those with scores < 4 delivered immediately. Reference standard(s): "Poor neonatal outcome," which included neonatal distress requiring admission to the NICU, endotracheal intubation, use of positive pressure oxygen for more than 6 hours, and persistent fetal circulation Dates: NR Location: Boston, MA Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 131 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 131 Inclusion criteria: Gestational age ≥ 41 completed weeks; singleton pregnancy; no congenital anomalies Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry: NR (gestational age ≥ 41 completed weeks required for entry into study) Dating criteria: U/S early in pregnancy or reliable menstrual dates and serial physical exams Parity: NR Bishop score: NR	1) Meconium aspiration 2) Admission to NICU 3) Seizure 4) 2 x 2 table 5) Predictive values	1) Meconium aspiration: 5/131 (4%) 2) Admission to NICU: 5/131 (4%) 3) Seizure: 1/131 (< 1%) 4) 2 x 2 table: Reference standard = Poor neonatal outcome Screening test = Biophysical Profile Score (BPS) Neonatal outcome poor / normal / Totals: BPS abn (< 6): 1 / 7 / 8 BPS nl (≥ 6): 6 / 117 / 123 Totals: 7 / 124 / 131 5) Predictive values: *Positive predictive values:* Total BPS: 14% Amniotic fluid volume: 17% Placental grading: 4% Fetal breathing movements: 5% Fetal tone/movements: 40% NST: 14% *Negative predictive values:* Total BPS: 94% Amniotic fluid volume: 95% Placental grading: 91% Fetal breathing movements: 94% Fetal tone/movements: 95% NST: 94%	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: + Test reliability/variability: - Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: - Testing protocol described: - Sample size: - Statistical tests: +

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Imai, Tani, Saito, et al., 2001	Design: Cohort study Test(s) studied: 1) Fetal fibronectin obtained from posterior vaginal fornix. Collected once between 29 and 35 weeks, then weekly from 36 weeks until parturition. 2) Cytokines Interleukin-1, beta, IL-6, IL8, and tumor necrosis factor alpha. Collected from endocervix at same intervals as above. Reference standard(s): Delivery within 7 days of sampling Dates: NR Location: Kanagawa, Japan Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 122 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 120 (2 excluded for no labor) Inclusion criteria: Singleton pregnancy; vertex presentation Exclusion criteria: Maternal or obstetric complications that might cause premature delivery, premature rupture of membranes, vaginal bleeding, or fetal anomalies Age: Mean, 30; range, 20-45 Race: NR Gestational age at entry: NR (gestational age between 29 and 35 weeks required for entry into study) Dating criteria: LMP, confirmed by ultrasound prior to 20 weeks Parity: 71% nulliparous Bishop score: NR	1) Fetal fibronectin 2) IL-1 beta 3) 2x2 tables	1) Fetal fibronectin: At threshold of > 50 ng/ml: Sensitivity: 90% Specificity: 51% Positive predictive value: 75% Negative predictive value: 75% 2) IL-1 beta: At threshold of 100 pg/ml: Sensitivity: 55% Specificity: 76% Positive predictive value: 79% Negative predictive value: 50% 3) 2 x 2 tables: <u>2 x 2 table 1:</u> Reference standard = Delivery within 7 days Screening test = Fetal Fibronectin (fFN) Time to delivery ≤ 7 days / > 7 days / Totals: fFN > 50: 120 / 39 / 159 fFN ≤ 50: 13 / 40 / 53 Totals: 133 / 79 / 212 <u>2 x 2 table 2:</u> Reference standard = Delivery within 7 days Screening test = IL-2 beta Time to delivery ≤ 7 days / > 7 days / Totals: IL-2 > 100: 73 / 19 / 92 IL-2 ≤ 100: 60 / 60 / 120 Totals: 133 / 79 / 212	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: + Test reliability/variability: - Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: - Reported sensitivity/specificity was reversed in tables and text of article; values from text used here. Any variations by gestational age within the 36-42 week gestational range not reported.

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Jazayeri, Heffron, Phillips, et al., 1999	Design: Case series (retrospective), concomitant controls Test(s) studied: 1) U/S measuring estimated fetal weight, abdominal circumference, biparietal diameter, and femur length Protocol: Measurements taken within 2 weeks of delivery Reference standard(s): 1) Macrosomia Dates: Jan-Dec 1996 Location: Tampa, FL Setting: University hospital Type(s) of providers: MFM Length of follow-up: None	No. of subjects at start: 168 (84 with macrosomic infants; 84 with nonmacrosomic infants) Dropouts: NA (retrospective study) Loss to follow-up: NA No. of subjects at end: 168 Inclusion criteria: Women with macrosomic infants (≥ 4000 g) and U/S within 2 weeks prior to delivery; these women compared with group of women with non-macrosomic infants and recent U/S Exclusion criteria: None specified Age (mean ± SD): Macrosomic, 25.9 ± 6; nonmacrosomic, 24.4 ± 5 Race: Macrosomic, 45% White, 25% Black, 30% Hispanic; non-macrosomic, 40% White, 30% Black, 30% Hispanic Gestational age at entry (mean ± SD): Macrosomic, 40.1 ± 1.5 weeks; nonmacrosomic, 37.1 ± 3.6 weeks (p = 0.001) Dating criteria: NR Gravidity (median, with range): Macrosomic, 3 ± 2; non-macrosomic, 2 ± 1 Bishop score: NR	1) Shoulder dystocia 2) C-section for fetal distress 3) 2 x 2 table 4) Other test performance results	1) Shoulder dystocia: Macrosomic: 13/84 (15%) Nonmacrosomic: 0/84 p = 0.001 In macrosomic newborns, labor induction was associated with a 22% rate of should dystocia, whereas augmentation and spontaneous labor were each associated with an 8% rate of shoulder dystocia (odds ratio, 3.4; 95% CI, 1.4 to 8.2; p < 0.01). In a multivariable model controlling for birthweight, induction alone was associated with shoulder dystocia (odds ratio, 2.97; 95% CI, 1.24 to 7.4; p < 0.015). 2) C-section for fetal distress: Macrosomic: 25/84 (30%) Nonmacrosomic: 19/84 (23%) p = not significant 3) 2 x 2 table: Reference standard = Macrosomia (≥ 4000 g) Screening test = U/S measurement of abdominal circumference (AC) Macrosomia yes / no / Totals: AC ≥ 35 cm: 75 / 6 / 81 AC < 35 cm: 9 / 78 / 87 Totals: 84 / 84 / 168 4) Other test performance results: Multiple regression analysis showed abdominal circumference to be the best predictor of birthweight in macrosomic infants.	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: + Dating criteria: - Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: - Sample size: - Statistical tests: +

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Knox, Huddleston, and Flowers, 1979	Design: RCT, allocation to group by last digit of hospital number Test(s) studied: 1) Amniocentesis (n = 90) Protocol: If no meconium discovered and fluid obtained, then amniocentesis repeated in 1 week. If meconium discovered or no fluid obtained, then labor induced. Labor induced with IV oxytocin, with direct FHR and intrauterine pressure monitoring. 2) Oxytocin challenge test (OCT) (n = 90) Protocol: Initial amniocentesis followed by OCT. If meconium present or no fluid discovered on amniocentesis, then labor induced. If OCT negative, the repeated in 1 week. If OCT positive, the labor induced. Reference standard(s): 1) Low birthweight 2) Neonatal morbidity 3) Perinatal death 4) C-sections 5) Apgar scores at 1 minute 6) Apgar scores at 5 minutes Dates: Aug 1975 - July 1976 Location: Birmingham, AL Setting: University hospital Type(s) of providers: Unspecified OB/GYN	No. of subjects at start: 187 Dropouts: 7 (excluded due to complications) Loss to follow-up: NA No. of subjects at end: 180 Inclusion criteria: Gestational age ≥ 42 weeks Exclusion criteria: Any obstetric complication Age: NR Race: NR Gestational age at entry: NR (gestational age ≥ 42 weeks required for entry into study) Dating criteria: Either a) reliable LMP confirmed by pelvic exam prior to 12 weeks, U/S at 20-30 weeks, or auscultation of unamplified fetal heart tones for at least 22 weeks; or b) if LMP unreliable, then 2 of above 3 assessments consistent with 42 weeks' gestation Parity: NR Bishop score: NR	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Low birthweight (< 10th percentile) 4) Neonatal morbidity 5) Perinatal death 6) Meconium 7) C-sections 8) Induction 9) Abnormal labor (prolonged latent phase, primary dysfunctional labor, secondary arrest of dilatation, or arrest of descent) 10) 2 x 2 tables 11) Other test performance results	1) Apgar score < 7 at 1 minute: Amniocentesis: 19/90 (21%) OCT: 12/90 (13%) p = not significant 2) Apgar score < 7 at 5 minutes: Amniocentesis: 6/90 (7%) OCT: 2/90 (2%) p = not significant 3) Low birthweight (< 10th percentile): Amniocentesis: 3/90 (3%) OCT: 4/90 (4%) p = not significant 4) Neonatal morbidity: Amniocentesis: 6/90 (7%) OCT: 7/90 (8%) p = not significant 5) Perinatal death: Amniocentesis: 3/90 (3%) OCT: 1/90 (1%) p = not significant 6) Meconium (overall only): On initial amniocentesis: 22% At delivery: 44% 7) C-sections: Amniocentesis: 11/90 (12%) OCT: 8/90 (9%) p = not significant 8) Induction: Amniocentesis: 29/90 (32%) OCT: 11/90 (12%) p < 0.005 9) Abnormal labor: Amniocentesis: 13/90 (14%) OCT: 12/90 (13%) p = not significant	QUALITY SCORES: TESTING Reference standard: + Randomized: + Method of randomization: - Verification bias: + Test reliability/variability: - Gestational age: - Dating criteria: + Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: + MANAGEMENT Randomized: + Method of randomization: - Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: - Dating criteria: + Bishop score: - *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Length of follow-up: None			10) 2 x 2 tables: 2 x 2 table 1: Reference standard = Low birthweight (< 10th percentile) Screening test = Meconium at initial amniocentesis	

	Low birthweight yes	no	Totals:
Meconium present	2	77	79
Meconium absent	5	96	101
Totals:	7	173	180

2 x 2 table 2:
Reference standard = Neonatal morbidity
Screening test = Meconium at initial amniocentesis

	Morbidity yes	no	Totals:
Meconium present	6	73	79
Meconium absent	7	94	101
Totals:	13	167	180

2 x 2 table 3:
Reference standard = Perinatal death
Screening test = Meconium at initial amniocentesis

	Death yes	no	Totals:
Meconium present	4	75	79
Meconium absent	0	101	101
Totals:	4	176	180

2 x 2 table 4:
Reference standard = C-sections
Screening test = Meconium at initial amniocentesis

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

C-section

	yes	no	Totals:
Meconium present	11	68	79
Meconium absent	8	93	101
Totals:	19	161	180

2 x 2 table 5:
Reference standard = Apgar score at 1 minute
Screening test = Meconium at initial amniocentesis

Apgar at 1 min

	< 7	≥ 7	Totals:
Meconium present	23	56	79
Meconium absent	8	93	101
Totals:	31	149	180

2 x 2 table 6:
Reference standard = Apgar score at 5 minutes
Screening test = Meconium at initial amniocentesis

Apgar at 5 min

	< 7	≥ 7	Totals:
Meconium present	8	71	79
Meconium absent	0	101	101
Totals:	8	172	180

11) Other test performance results:
In subset of patients with meconium present, there were no significant differences between the two groups for any outcome.

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Leveno, Quirk, Cunningham, et al., 1984	Design: Cohort study (prospective) Test(s) studied: 1) Amniotic fluid volume (AFV) assessment Protocol: AFV assessed weekly. Oligohydramnios defined as two or fewer 1-cm pockets of amniotic fluid. If any of the following occurred, then labor was induced using oxytocin followed by amniotomy: a) certain completion of 43 weeks' gestation; b) absence of amniotic fluid on physical exam; c) markedly diminished fetal activity; or d) development of pregnancy-induced hypertension. Intrapartum electronic FHR monitoring used. Reference standard(s): 1) C-section for fetal distress 2) Small for gestational age 3) Stillbirth or meconium aspiration Dates: July 1980 - July 1982 Location: Dallas, TX Setting: University hospital Type(s) of providers: MFM Length of follow-up: None	No. of subjects at start: 727 (of whom 213 underwent U/S assessment of AFV) Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 727 Inclusion criteria: Gestational age ≥ 41 completed weeks Exclusion criteria: Obstetric or medical complications Age: 55% were age 20-30 Race: 39% White, 39% Black, 22% Hispanic Gestational age at entry: 42-43 weeks (certain): 16% 43-44 weeks (certain): 8% > 44 weeks (certain): 1% Uncertain prolonged pregnancy: 75% Dating criteria: LMP corroborated by a) fetal heart auscultation between 17 and 20 weeks; *or* b) fundal height measurements between 20 and 30 weeks; *or* c) U/S before 26 weeks Parity: "Approximately half" were nulliparous Bishop score: NR	1) C-sections 2) 2 x 2 tables	1) C-sections: Overall: 196/727 (27%) For cephalopelvic disproportion: 114/727 (16%) For fetal distress:* 59/727 (8%) For abnormal presentation: 16/727 (2%) For other reasons: 7/727 (1%) *"Fetal distress" diagnosed when one or more of the following were identified on intrapartum FHR monitoring: a) repetitive late decelerations; b) severe variable decelerations of < 60 bpm for ≥ 1 minute; c) prolonged decelerations lasting ≥ 2 minutes; or d) unexplained abnormal baseline heart rate or diminished beat-to-beat variability, especially when either accompanied by meconium staining. 2) 2 x 2 tables (for women undergoing AFV assessment only, n = 213) <u>2 x 2 table 1:</u> Reference standard = C-section for fetal distress (as defined above) Screening test = Amniotic fluid volume (AFV) assessment C-section yes / no / Totals: AFV decreased: 11 / 73 / 84 AFV normal: 7 / 122 / 129 Totals: 18 / 195 / 213 <u>2 x 2 table 2:</u> Reference standard = Small for gestational age (SGA) Screening test = AFV	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: + Test reliability/variability: + Gestational age: + Dating criteria: + Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: +

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

	SGA yes	SGA no	Totals:
AFV decreased	8	76	84
AFV normal	8	121	129
Totals:	16	197	213

2 x 2 table 3:
Reference standard = Stillbirth or meconium aspiration
Screening test = AFV

	Stillbirth/ meconium yes	Stillbirth/ meconium no	Totals:
AFV decreased	2	82	84
AFV normal	0	129	129
Totals:	2	211	213

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Monaghan, O'Herlihy, and Boylan, 1987	Design: Cohort study (not specified if prospective or retrospective) Test(s) studied: 1) Ultrasound used to measure deepest amniotic fluid pool and to grade placental echogenic changes (n = 200) Protocol: U/S scans performed every 3-5 days beginning at 42 weeks. Used to measure deepest vertical amniotic fluid pool. If no pool exceeded 30 mm, then oligohydramnios diagnosed and labor induced. U/S also used to grade echogenic characteristics of placenta from 0 (homogeneous placenta with smooth chorionic plate) to III (placenta completely divided into compartments by indentation of the chorionic plate extending all the way to the basal layer). Placental grading not used to make management decisions. Reference standard(s): 1) Fetal acidosis 2) C-section for fetal distress 3) Low birthweight 4) Admission to NICU 5) Perinatal death Dates: NR Location: Dublin, Ireland Setting: Unspecified hospital Type(s) of providers:	No. of subjects at start: 225 Dropouts: 25 (excluded because of uncertain gestational age) Loss to follow-up: NA No. of subjects at end: 200 Inclusion criteria: Gestational age ≥ 42 weeks; singleton pregnancy Exclusion criteria: Uncertain gestational age Age: NR Race: NR Gestational age at entry: NR (gestational age ≥ 42 weeks required for entry into study) Dating criteria: Certain LMP or early U/S Parity: 41% primiparous Bishop score: NR	1) Fetal acidosis (pH < 7.25) 2) Low birthweight (< 10th percentile) 3) Admission to NICU 4) Perinatal death 5) Inductions 6) C-sections 7) 2 x 2 tables 8) Other test performance results	1) Fetal acidosis: 13/200 (7%) 2) Low birthweight: 23/200 (12%) 3) Admission to NICU: 18/200 (9%) 4) Perinatal death: 2/200 (1%) 5) Inductions: 69/200 (35%) Labor induced in 32 cases because of oligohydramnios, and in 37 cases with favorable cervical status and normal amniotic fluid estimates. 6) C-sections: Overall: 12/200 (6%) For fetal distress: 3/200 (2%) 7) 2 x 2 tables: <u>2 x 2 Table 1:</u> Reference standard = Fetal acidosis (pH < 7.25) Screening test = Amniotic fluid index (AFI) ("low" if no pool exceeded 30 mm) Fetal acidosis yes / no / Totals: AFI low: 3 / 29 / 32 AFI normal: 10 / 158 / 168 Totals: 13 / 187 / 200 <u>2 x 2 Table 2:</u> Reference standard = C-section for fetal distress Screening test = AFI (as above) C-section yes / no / Totals: AFI low: 1 / 31 / 32 AFI normal: 2 / 166 / 168 Totals: 3 / 197 / 200	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: + Test reliability/variability: - Gestational age: + Dating criteria: + Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: +

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Unspecified OB/GYN Length of follow-up: None			2 x 2 Table 3: Reference standard = Low birthweight (< 10th percentile) Screening test = AFI (as above) Low birthweight yes / no / Totals: AFI low 11 / 21 / 32 AFI normal 12 / 156 / 168 Totals: 23 / 177 / 200 2 x 2 Table 4: Reference standard = Admission to NICU Screening test = AFI (as above) NICU admission yes / no / Totals: AFI low 3 / 29 / 32 AFI normal 15 / 153 / 168 Totals: 18 / 182 / 200 2 x 2 Table 5: Reference standard = Perinatal death Screening test = AFI (as above) Perinatal death yes / no / Totals: AFI low 0 / 32 / 32 AFI normal 2 / 166 / 168 Totals: 2 / 198 / 200 8) Other test performance results: Ultimate placental grading was associated with an increased incidence of C-section. The increased incidence associated with grade III placenta was related to mothers with coincident oligohydramnios. The frequency of meconium staining and no amniotic fluid after amniotomy was higher in patients with oligohydramnios. There were no differences in acidosis or	*(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				NICU admission between pregnancies with normal versus reduced amniotic fluid, or grade 1-11 versus grade III placentas. The incidence of low birthweight was significantly higher in patients with oligohydramnios than in patients with grade III placentas.	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Montan and Malcus, 1995	Design: Cohort study (prospective) Test(s) studied: 1) Amniotic fluid index (AFI) and FHR pattern Protocol: AFI and FHR pattern measured at 2-day intervals from 42 weeks until delivery. Labor induced (by oxytocin or artificial rupture of the membranes) for abnormal fetal or maternal findings. Reference standard(s): 1) C-section 2) Apgar < 7 at 1 minute 3) Apgar < 7 at 5 minutes Dates: 1992-93 Location: Ängelholm, Sweden Setting: Community hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 116 women delivered at ≥ 42 weeks gestation; 88 of them had AFI measured at least once before onset of labor Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 116 Inclusion criteria: Gestational age ≥ 42 completed weeks Exclusion criteria: None specified Age (mean, with range): 28 (17-46) Race: NR Gestational age at entry: NR (gestational age required to be ≥ 42 completed weeks for entry into study) Dating criteria: U/S (biparietal diameter and femur length) in weeks 16-19 Parity: 49% primigravida Bishop score: NR	1) 2 x 2 tables 2) Other test performance results	1) 2 x 2 tables 2 x 2 table 1: Reference standard = C-section Screening test = AFI C-section yes / no / Totals: AFI < 5 cm: 1 / 10 / 11 AFI ≥ 5 cm: 11 / 66 / 77 No AFI: 7 / 21 / 28 Totals: 19 / 97 / 116 2 x 2 table 2: Reference standard = Apgar score at 1 minute Screening test = AFI Apgar at 1 min < 7 / ≥ 8 / Totals: AFI < 5 cm: 0 / 11 / 11 AFI ≥ 5 cm: 3 / 74 / 77 No AFI: 1 / 27 / 28 Totals: 4 / 112 / 116 2 x 2 table 3: Reference standard = Apgar score at 5 minutes Screening test = AFI Apgar at 5 min < 7 / ≥ 8 / Totals: AFI < 5 cm: 0 / 11 / 11 AFI ≥ 5 cm: 2 / 75 / 77 No AFI: 0 / 28 / 28 Totals: 2 / 114 / 116 2) Other test performance results: There was no association between low AFI (< 5 cm) and signs of fetal distress expressed as abnormal FHR pattern, meconium staining, Apgar scores < 7, or C-section.	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: + Test reliability/variability: + Gestational age: + Dating criteria: + Other risk factors absent: - Similar to likely pt pop: - Testing protocol described: + Sample size: - Statistical tests: + The definition of low AFI used in this study (< 5 cm) is more liberal than that used in many studies (3 cm or 1 cm) and may explain the lack of association between low AFI and fetal compromise reported here.

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Mouw, Egberts, Kragt, et al., 1998	Design: Case series (prospective), no controls Test(s) studied: 1) Fetal fibronectin concentration (fFN) Protocol: Fetal fibronectin concentration measured in cervicovaginal secretions obtained in sterile speculum examination at 41 weeks. Concentrations of < 50 ng/ml were interpreted as negative, ≥ 50 ng/ml as positive. Pregnancies were managed expectantly, and induction was performed only for obstetric "or sometimes psychological" reasons. Reference standard(s): 1) Birth within 3 days of fFN testing Dates: NR Location: Leiden and Voorburg, The Netherlands Setting: 2 university hospitals Type(s) of providers: General OB/GYN Length of follow-up: None	No. of subjects at start: 80 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 80 Inclusion criteria: Gestational age ≥ 41 weeks Exclusion criteria: In labor; clinical evidence of ruptured membranes Age (mean ± SD): 31 ± 6 Race: NR Gestational age at entry: Range, 287-304 days Dating criteria: NR Parity (mean ± SD): 1 ± 1 Bishop score: NR	1) 2 x 2 table 2) Other test performance results	1) 2 x 2 table: Reference standard = Birth within 3 days of fFN testing Screening test = fFN Birth within 3 days yes / no / Totals: fFN ≥ 50 ng/ml: 30 / 15 / 45 fFN< 50 ng/ml: 12 / 27 / 39 Totals: 42 / 42 / 84 2) Other test performance results: A positive fFN test (≥ 50 ng/ml) had sensitivity of 0.71 (95% CI, 0.58 to 0.86) and specificity of 0.64 (95% CI, 0.48 to 0.78) for predicting birth within 3 days. The change from negative to positive fFN values often occurred between 1 and 4 days before birth in women with a spontaneous onset of labor. The mean interval between positive test and birth was 2.5 ± 2.5 days (range, 0-11). fFN was moderately correlated with Bishop score. Bishop score > 5 had sensitivity 0.67 (95% CI, 0.48 to 0.82) and specificity 0.77 (95% CI, 0.54 to 0.92) for predicting birth within 3 days. (Only 74% of study participants had Bishop scores recorded.)	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: + Test reliability/variability: + Gestational age: + Dating criteria: - Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: + Statistical tests: + Sensitivity/specificity results include some repeat tests.

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
O'Reilly-Green and Divon, 1996	Design: Cohort study (retrospective) Test(s) studied: 1) Sonographic estimate of fetal weight (EFW) plus nonstress test (NST) and amniotic fluid index (AFI) Protocol: Sonographic EFW done at initial appointment. NST and AFI performed twice weekly. If AFI ≤ 5 cm, then patient delivered within 24 hours, even if all other testing parameters were normal. Reference standard(s): 1) Apgar score at 1 minute 2) Apgar score at 5 minutes 3) Any complication Dates: July 1991- Sep 1992 Location: Bronx, NY Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 449 Dropouts: NA (retrospective study) Loss to follow-up: NA No. of subjects at end: 449 Inclusion criteria: Prolonged pregnancy (defined as 1 or more weeks beyond expected date of delivery) Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry: NR Dating criteria: Nagle's rule or sonographic criteria Parity: NR Bishop score: NR	1) Apgar score < 8 at 1 minute 2) Apgar score < 9 at 5 minutes 3) 2 x 2 tables 4) Other test performance results	1) Apgar score < 8 at 1 minute: 66/449 (15%) 2) Apgar score < 9 at 5 minutes: 24/449 (5%) 3) 2 x 2 tables: 2 x 2 Table 1: Reference standard = Apgar score at 1 minute Screening test = Amniotic fluid index (AFI) Apgar at 1 min < 8 / ≥ 8 / Totals: AFI ≤ 5: 5 / 45 / 50 AFI > 5: 61 / 337 / 398 Totals: 66 / 382 / 448 2 x 2 Table 2: Reference standard = Apgar score at 5 minutes Screening test = Amniotic fluid index (AFI) Apgar at 5 min < 9 / ≥ 9 / Totals: AFI ≤ 5: 2 / 48 / 50 AFI > 5: 22 / 376 / 398 Totals: 24 / 424 / 448 2 x 2 Table 3: Reference standard = Any complication Screening test = Amniotic fluid index (AFI) Complication yes / no / Totals: AFI ≤ 5: 4 / 46 / 50 AFI > 5: 25 / 372 / 397 Totals: 29 / 418 / 447 4) Other test performance results: Additional analyses showed significant association between AFI ≤ 5 and clinical oligohydramnios.	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: - Dating criteria: - Other risk factors absent: - Similar to likely pt pop: - Testing protocol described: + Sample size: - Statistical tests: - Same population as in O'Reilly-Green and Divon, 1997, below.

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
O'Reilly-Green and Divon, 1997	Design: Cohort study (retrospective) Test(s) studied: 1) Sonographic estimate of fetal weight Protocol: Estimate made ≤ 21 days before admission (≤ 22 days before delivery). Estimated fetal weight (EFW) calculated using formulas at the discretion of the clinician interpreting the study. An adjusted EFW was calculated by adding 12.7 g to the EFW for each day that elapsed between the sonographic measurements and delivery. Reference standard(s): 1) Actual birthweight Dates: July 1991 - Sep 1992 Location: Bronx, NY Setting: University hospital Type(s) of providers: General OB/GYN Length of follow-up: None	No. of subjects at start: 445 Dropouts: NA (retrospective study) Loss to follow-up: NA No. of subjects at end: 445 Inclusion criteria: Prolonged pregnancy (defined as 4 or more days beyond expected date of delivery) Exclusion criteria: Diabetes Age: NR Race: NR Gestational age at entry (mean ± SD): 291 ± 6.7 days Dating criteria: Naegele's rule or sonographic criteria Parity: NR Bishop score: NR	1) 2 x 2 tables 2) Other test performance results	1) 2 x 2 tables: 2 x 2 Table 1: Reference standard = Actual birthweight Screening test = Estimated fetal weight (EFW) Birthweight ≥ 4000 g / < 4000 g / Totals: EFW ≥ 3711 g: 91 / 94 / 185 EFW < 3711 g: 16 / 244 / 260 Totals: 107 / 338 / 445 2 x 2 Table 2: Reference standard = Actual birthweight Screening test = EFW Birthweight ≥ 4500 g / < 4500 g / Totals: EFW ≥ 4192 g: 15 / 35 / 50 EFW < 4192 g: 3 / 392 / 395 Totals: 18 / 427 / 445 2 x 2 Table 3: Reference standard = Actual birthweight Screening test = EFW Birthweight ≥ 4000 g / < 4000 g / Totals: EFW ≥ 4000 g: 60 / 29 / 89 EFW < 4000 g: 47 / 309 / 356 Totals: 107 / 338 / 445 2 x 2 Table 4: Reference standard = Actual birthweight Screening test = EFW	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: + Test reliability/variability: + Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: + Testing protocol described: + Sample size: + Statistical tests: + Same population as in O'Reilly-Green and Divon, 1996, above.

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

	Birthweight ≥ 4500 g	Birthweight < 4500 g	Totals:
EFW ≥ 4500 g	4	5	9
EFW < 4500 g	14	422	436
Totals:	18	427	445

2) Other test performance results:
EFW ≥ 3711 g had sensitivity 0.85 and specificity 0.72 for predicting birthweight ≥ 4000 g.

EFW ≥ 4000 g had sensitivity 0.56 and specificity 0.91 for predicting birthweight ≥ 4000 g.

The area under ROC curve for EFW within 4 days of delivery as a predictor of birthweight ≥ 4000 g was 0.85; for 5-22 days, 0.85; and for 0-22 days, 0.85.

EFW ≥ 4192 g had sensitivity 0.83 and specificity 0.92 for predicting birthweight ≥ 4500 g.

EFW ≥ 4500 g had sensitivity 0.22 and specificity 0.99 for predicting birthweight ≥ 4500 g.

The area under ROC curve for EFW within 4 days of delivery as a predictor of birthweight ≥ 4500 g was 0.93; for 5-22 days, 0.95; and for 0-22 days, 0.95.

The area under ROC curve for the adjusted EFW within 4 days of delivery as a predictor of birthweight ≥ 4500 g was 0.93; for 5-22 days, 0.95; and for 0-22 days, 0.95.

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Phelan, Platt, Yeh, et al., 1984	Design: Case series (retrospective), no controls Test(s) studied: 1) Nonstress test (NST) (n = 239) Protocol: Last NST conducted within 7 days of delivery. NST considered reactive if ≥ 2 FHR accelerations of > 15 bpm, lasting 15 seconds, in a 20-min period. Reactive NSTs repeated in a week (or sooner if serum estriol was low). NST considered nonreactive if there were not 2 acceptable FHR accelerations in any 20-min period of observation totaling 40 minutes. If test nonreactive, then patient re-tested in afternoon. If afternoon test nonreactive, then CST performed (or, if CST contraindicated, then biophysical profile done). If CST negative, then repeated in 24 hours. Reference standard(s): 1) C-section for fetal distress 2) Meconium aspiration 3) Apgar score at 1 minute 4) Apgar score at 5 minutes 5) Macrosomia 6) Post-maturity syndrome Dates: July 1980 - June 1981 Location: Los Angeles, CA Setting: University hospital Type(s) of providers: General OB/GYN; specially trained antepartum nurses	No. of subjects at start: 239 Dropouts: NA (retrospective analysis) Loss to follow-up: NA No. of subjects at end: 239 Inclusion criteria: Post-dates (> 294 days); underwent NST within 7 days of delivery Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry: NR; gestational age > 294 days required for inclusion in study Dating criteria: LMP Parity: NR Bishop score: NR	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Meconium staining 4) Meconium aspiration 5) Macrosomia (birthweight ≥ 4000 g) 6) Post-maturity syndrome 7) C-sections 8) 2 x 2 tables 9) Other test performance results	1) Apgar score < 7 at 1 minute: 47/239 (20%) 2) Apgar score < 7 at 5 minutes: 6/239 (3%) 3) Meconium staining: 99/239 (41%) 4) Meconium aspiration: 19/239 (8%) 5) Macrosomia (birthweight ≥ 4000 g): 52/239 (22%) 6) Post-maturity syndrome: 40/239 (17%) 7) C-sections: Overall: 42/239 (18%) For fetal distress: 13/239 (5%) 8) 2 x 2 tables: <u>2 x 2 table 1:</u> Reference standard = C-section for fetal distress Screening test = Nonstress test (NST) C-section yes / no / Totals: NST nonreactive 4 / 28 / 32 NST reactive 9 / 198 / 207 Totals: 13 / 226 / 239 <u>2 x 2 table 2:</u> Reference standard = Meconium aspiration Screening test = NST Meconium aspiration yes / no / Totals: NST nonreactive 5 / 27 / 32	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: + Dating criteria: - Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: - Same patient population as Phelan, Platt, Yeh, et al. 1985, below. *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Length of follow-up: None			NST reactive 14 193 207 Totals: 19 220 239 2 x 2 table 3: Reference standard = Apgar score at 1 minute Screening test = NST Apgar at 1 min: < 7 / ≥ 7 / Totals: NST nonreactive 11 21 32 NST reactive 36 171 207 Totals: 47 192 239 2 x 2 table 4: Reference standard = Apgar score at 5 minutes Screening test = NST Apgar at 5 min: < 7 / ≥ 7 / Totals: NST nonreactive 2 30 32 NST reactive 4 203 207 Totals: 6 233 239 2 x 2 table 5: Reference standard = Macrosomia (birthweight ≥ 4000 g) Screening test = NST Macrosomia: yes / no / Totals: NST nonreactive 4 28 32 NST reactive 48 159 207 Totals: 52 187 239	

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				2 x 2 table 6: Reference standard = Post-maturity syndrome Screening test = NST Post-maturity yes / no / Totals: NST nonreactive: 4 / 28 / 32 NST reactive: 36 / 171 / 207 Totals: 40 / 199 / 239 9) Other test performance results: Among patients with reactive NSTs, those with decelerations had significant increases in C-sections for fetal distress, meconium passage, and Apgar scores < 7 at 5 minutes compared to those without decelerations.	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Phelan, Platt, Yeh, et al., 1985	Design: Case series (retrospective), no controls Test(s) studied: 1) Nonstress test (NST), biophysical profile, and amniotic fluid volume (AFV) Protocol: Testing schedule not described (though referenced). Patients with FHR bradycardia revealed on the NST were evaluated for delivery. AFV considered "adequate" if largest pocket > 1 cm in vertical diameter; "decreased" if largest pocket ≤ 1 cm; and "adequate, but decreased" if largest pocket > 1 cm, but overall impression of sonographer was that fluid was decreased. Reference standard(s): 1) C-section for fetal distress 2) Apgar score at 1 minute 3) Apgar score at 5 minutes 4) Birthweight Dates: July 1980 - June 1981 Location: Los Angeles, CA Setting: University hospital Type(s) of providers: General OB/GYN; specially trained antepartum nurses Length of follow-up: None	No. of subjects at start: 236 Dropouts: NA (retrospective study) Loss to follow-up: NA No. of subjects at end: 236 Inclusion criteria: Post-dates; underwent biophysical testing within 7 days of delivery Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry: NR Dating criteria: NR Parity: NR Bishop score: NR	1) Macrosomia (birthweight > 4000 g) 2) Apgar score < 7 at 1 minute 3) Apgar score < 7 at 5 minutes 4) Post-maturity syndrome 5) Meconium staining 6) Meconium aspiration 7) Deceleration or bradycardia 8) Fetal death 9) C-sections 10) 2 x 2 tables	1) Macrosomia (birthweight > 4000 g): 52/236 (22%) 2) Apgar score < 7 at 1 minute: 49/236 (21%) 3) Apgar score < 7 at 5 minutes: 8/236 (3%) 4) Post-maturity syndrome: 40/236 (17%) 5) Meconium staining: 99/236 (42%) 6) Meconium aspiration: 19/236 (8%) 7) Deceleration or bradycardia: 62/236 (26%) 8) Fetal death: 2/236 (< 1%) 9) C-sections: Overall: 45/236 (19%) For fetal distress: 13/236 (6%) 10) 2 x 2 tables: 2 x 2 table 1: Reference standard = C-section for fetal distress Screening test = Amniotic fluid volume (AFV) C-section yes / no / Totals: AFV decreased: 3 / 4 / 7 AFV adequate/ decreased: 6 / 32 / 38 AFV adequate: 4 / 187 / 191 Totals: 13 / 223 / 236	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: + Gestational age: - Dating criteria: - Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: + Same patient population as in Phelan, Platt, Yeh, et al., 1984, above.

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				2 x 2 table 2: Reference standard = Apgar score at 1 minute Screening test = AFV Apgar at 1 min: < 7 / ≥ 7 / Totals: AFV decreased: 6 / 1 / 7 AFV adequate/ decreased: 12 / 26 / 38 AFV adequate: 31 / 160 / 191 Totals: 49 / 187 / 236 2 x 2 table 3: Reference standard = Apgar score at 5 minutes Screening test = AFV Apgar at 5 min: < 7 / ≥ 7 / Totals: AFV decreased: 2 / 5 / 7 AFV adequate/ decreased: 1 / 37 / 38 AFV adequate: 5 / 186 / 191 Totals: 8 / 228 / 236 2 x 2 table 4: Reference standard = Birthweight Screening test = AFV Birthweight: > 4000 g / ≤ 4000 g / Totals: AFV decreased: 0 / 7 / 7 AFV adequate/ decreased: 6 / 32 / 38 AFV adequate: 46 / 145 / 191 Totals: 52 / 184 / 236	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Pollack, Hauer-Pollack, and Divon, 1992	Design: Case series (retrospective), no controls Test(s) studied: 1) Ultrasound examination to estimate fetal weight Protocol: Exam performed within 1 week of delivery. Estimate of fetal weight based on biparietal diameter, abdominal circumference, and femur length. Reference standard(s): 1) Macrosomia (defined using two different thresholds) Dates: Jan 1989 - Sep 1990 Location: Bronx, NY Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 519 Dropouts: NA (retrospective study) Loss to follow-up: NA No. of subjects at end: 519 Inclusion criteria: Gestational age ≥ 41 weeks; singleton pregnancy; U/S estimation of fetal weight within 1 week of delivery Exclusion criteria: Any complications of pregnancy Age: NR Race: NR Gestational age at entry: NR; gestational age ≥ 41 weeks required for inclusion in study Dating criteria: LMP and early U/S, when available; U/S dates preferred when there was a discrepancy of > 10 days between menstrual dates and U/S Parity: NR Bishop score: NR	1) 2 x 2 tables 2) Other test performance results	1) 2 x 2 tables: <u>2 x 2 table 1:</u> Reference standard = Macrosomia (defined as birthweight > 4000 g) Screening test = Estimated fetal weight (EFW) Birthweight \> 4000 g ≤ 4000 g Totals: EFW ≥ 4000 g 67 36 103 EFW < 4000 g 52 364 416 Totals: 119 400 519 <u>2 x 2 table 2:</u> Reference standard = Macrosomia (defined as birthweight > 4500 g) Screening test = EFW Birthweight \> 4500 g ≤ 4500 g Totals: EFW ≥ 4500 g 3 6 9 EFW < 4500 g 18 492 510 Totals: 21 498 519 2) Other test performance results: *EFW > 4000 g as a predictor of macrosomia (> 4000 g):* Sensitivity: 0.56 Specificity: 0.91 Positive predictive value: 0.64 Negative predictive value: 0.87 *EFW > 4500 g as a predictor of macrosomia (> 4500 g):* Sensitivity: 0.15 Specificity: 0.99 Positive predictive value: 0.81 Negative predictive value: 0.80	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: + Gestational age: - Dating criteria: - Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: Sample size: - Statistical tests: +

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Ramreker-singh-White, Farkas, Chard, et al., 1993	Design: Case series, no controls Test(s) studied: 1) Blood pressure, urine analysis, maternal weight, fetal movements, cardiotocography, and Doppler U/S velocimetry of utero-placental and umbilical blood flow (n = 167) Protocol: Above-mentioned tests performed twice weekly Reference standard(s): 1) Meconium staining Dates: 1991 Location: London, UK Setting: Unspecified hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 167 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 167 Inclusion criteria: Gestational age ≥ 280 days; uncomplicated pregnancy Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry: NR (gestational age ≥ 280 days required for entry into study) Dating criteria: LMP and U/S at 16 weeks Parity: NR Bishop score: NR	1) Meconium staining 2) Fetal distress (defined as a cardiotocographic abnormality significant enough to lead to operative delivery) 3) Stillbirth 4) 2 x 2 table 5) Other test performance results	1) Meconium staining: 15/167 (9%) 2) Fetal distress: 16/167 (10%) 3) Stillbirth: 1/167 (< 1%) 4) 2 x 2 table: Reference standard = Meconium staining Screening test = Fetal distress (defined at left) |	Meconium yes	no	Totals:
---	---	---	---
Fetal distress	5	11	16
No fetal distress	10	141	151
Totals:	15	152	167	 5) Other test performance results: There were no differences in mean Doppler indices (resistance index for right and left arcuate arteries, resistance and pulsatility indices for umbilical artery) between the 16 women with fetal distress and the remaining 151 women. No quantitative data reported.	QUALITY SCORE: Reference standard: - Randomized: - Method of randomization: NA Verification bias: + Test reliability/variability: - Gestational age: - Dating criteria: + Other risk factors absent: + Similar to likely pt pop: - Testing protocol described: + Sample size: - Statistical tests: -

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Rayburn, Motley, Stempel, et al., 1982	Design: Cohort study (prospective) Test(s) studied: 1) Nonstress test (NST) + fetal movement charting + urine estrogen-to-creatinine ratio. Protocol: Above-mentioned tests performed semi-weekly or weekly. If NST reactive (≥ 2 adequate accelerations of baseline FHR during a 20- to 40-minute period), then repeated on the next visit. If NST nonreactive, then test either repeated or a CST given the same day. Reference standard(s): 1) Post-maturity syndrome Dates: July 1979 - Apr 1981 Location: Columbus, OH Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 147 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 147 Inclusion criteria: Gestational age ≥ 42 weeks; scheduled to undergo NST Exclusion criteria: None specified Age: 20% ≤ 19; 69% 20-29; 11% ≥ 30 Race: 66% White, 34% Black Gestational age at entry: 42-43 weeks: 69% 43-44 weeks: 22% ≥ 44 weeks: 9% Dating criteria: LMP + either physical exam before 12th week or U/S before 20th week Parity: 46% primiparous Bishop score: NR Other: Cervical dilation: > 2 cm: 24% ≤ 2 cm: 76%	1) Post-maturity syndrome 2) Admission to NICU 3) Meconium aspiration 4) Birth asphyxia 5) Death 6) 2 x 2 tables 7) Other test performance results	1) Post-maturity syndrome: 32/147 (22%) 2) Admission to NICU: 7/147 (5%) 3) Meconium aspiration: 3/147 (2%) 4) Birth asphyxia: 1/147 (1%) 5) Death: 1/147 (1%) 6) 2 x 2 tables: <u>2 x 2 table 1:</u> Reference standard = Post-maturity syndrome Screening test = Antepartum FHR monitoring |	Post-maturity yes	no	Totals:
---	---	---	---
FHR abnormal	3	0	3
FHR normal	29	115	144
Totals:	32	115	147 | <u>2 x 2 table 2:</u> Reference standard = Post-maturity syndrome Screening test = Urine estrogen-to-creatinine ratio (E:C) |	Post-maturity yes	no	Totals:
---	---	---	---
E:C subnormal	12	0	12
E:C normal	3	50	53
Totals:	15	50	65	 <u>2 x 2 table 3:</u> Reference standard = Post-maturity syndrome Screening test = Fetal movement (FM) charting	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: + Placenta grading was not possible in 70/147 cases (48%) because ultrasonic visualization was too poor.

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

	Post-maturity yes	no	Totals:
FM inactive	0	0	0
FM active	32	115	147
Totals:	32	115	147

2 x 2 table 4:
Reference standard = Post-maturity syndrome
Screening test = Amniotic fluid volume (AFV)

	Post-maturity yes	no	Totals:
AFV adequate	24	5	29
AFV pockets	5	48	53
Oligo-hydramnios	3	62	65
Totals:	32	115	147

2 x 2 table 5:
Reference standard = Post-maturity syndrome
Screening test = Fetal motion (FM) on U/S

	Post-maturity yes	no	Totals:
FM absent	11	6	17
FM present	21	109	130
Totals:	32	115	147

2 x 2 table 6:
Reference standard = Post-maturity syndrome
Screening test = Fetal breathing (FB) on U/S

	Post-maturity yes	no	Totals:
FB absent	15	32	47
FB present	17	83	100
Totals:	32	115	147

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				7) Other test performance results: For predicting post-maturity syndrome: Test — Sensitivity — Specificity Oligo-hydramnios — 75% — 96% FHR testing — 9% — 100% Fetal movement charting — 0% — 100% E:C ratio — 80% — 100% Grade 3 placenta — 100% — 11% Gross fetal body motion — 34% — 95% Fetal breathing — 47% — 72%	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Sarkar and Duthie, 1997	Design: Cohort study (retrospective) Test(s) studied: 1) Cardiotocography and amniotic fluid index (AFI) (n = 184) Protocol: Cardiotocography and AFI performed twice weekly. Protocol not specified; presumably if AFI reduced, labor induced and continuous FHR monitoring used. Reference standard(s): 1) Birthweight 2) Apgar score at 5 minutes 3) Intubation 4) Admission to NICU 5) Emergency C-section Dates: Jan 1993 - Dec 1994 Location: Chester, UK Setting: Unspecified hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 184 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 184 Inclusion criteria: Gestational age ≥ 42 completed weeks; uncomplicated singleton pregnancy Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry: NR (gestational age ≥ 42 weeks required for entry into study) Dating criteria: LMP and U/S dates within 10 days of one another Parity: NR Bishop score: NR	1) Meconium staining 2) Low birthweight (< 5th percentile) 3) Apgar score < 7 at 5 minutes 4) Intubation 5) Admission to NICU 6) Abnormal FHR tracings 7) Emergency C-section (for fetal distress) 8) 2 x 2 tables	1) Meconium staining: 18/184 (9.8%) 2) Low birthweight (< 5th percentile): 2/184 (1%) 3) Apgar score < 7 at 5 minutes: 9/184 (4.9%) 4) Intubation: 5/184 (2.7%) 5) Admission to NICU: 1/184 (0.5%) 6) Abnormal FHR tracings: 47/184 (25.5%) 7) Emergency C-section (for fetal distress): 36/184 (19.6%) 8) 2 x 2 tables: 2 x 2 table 1: Reference standard = Birthweight ("low" defined as < 5th percentile) Screening test = AFI Birthweight low / normal / Totals: AFI decreased: 2 / 16 / 18 AFI normal: 0 / 166 / 166 Totals: 2 / 182 / 184 2 x 2 table 2: Reference standard = Apgar score at 5 minutes Screening test = AFI Apgar at 5 min < 7 / ≥ 7 / Totals: AFI decreased: 0 / 18 / 18 AFI normal: 9 / 157 / 166 Totals: 9 / 175 / 184	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: - Dating criteria: + Other risk factors absent: + Similar to likely pt pop: - Testing protocol described: + Sample size: - Statistical tests: +

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				2 x 2 table 3: Reference standard = Intubation Screening test = AFI Intubation yes / no / Totals: AFI decreased: 0 / 18 / 18 AFI normal: 5 / 161 / 166 Totals: 5 / 179 / 184 2 x 2 table 4: Reference standard = Admission to NICU Screening test = AFI Admission to NICU yes / no / Totals: AFI decreased: 1 / 17 / 18 AFI normal: 0 / 166 / 166 Totals: 1 / 183 / 184 2 x 2 table 5: Reference standard = Emergency C-section Screening test = AFI C-section yes / no / Totals: AFI decreased: 6 / 12 / 18 AFI normal: 30 / 136 / 166 Totals: 36 / 148 / 184	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Schreyer, Bar-Natan, Sherman, et al., 1991	Design: Case series (prospective), no controls Test(s) studied: 1) Fetal breathing movements (n = 65) Protocol: Fetal breathing movements were measured by U/S immediately before elective induction for reactive NST. Fetal breathing was considered to be present (+) when sustained for ≥ 20 seconds, and absent (-) when no sustained movement could be detected over a 45-minute period. Bishop score was assessed. Patients with Bishop score 0-2 were eliminated from the study and treated expectantly or by intracervical PGE_2 gel application. Labor was induced with oxytocin at 2 mIU/min, increasing by 1 mIU/min every 30 minutes until 3 contractions per 10 minutes. When cervix effaced and dilated 2-3 cm, membranes were artificially ruptured and internal cardiotocography initiated. Reference standard(s): 1) C-section for fetal distress 2) Apgar score at 5 minutes 3) Macrosomia Dates: June 1988 - June 1989 Location: Tel Aviv, Israel Setting: University hospital	No. of subjects at start: NR Dropouts: NR Loss to follow-up: NA No. of subjects at end: 65 Inclusion criteria: Gestational age 287-294 days Exclusion criteria: Pregnancy-induced hypertension; diabetes mellitus; previous C-section; IUGR; estimated fetal weight > 4300 g; malpresentation Age (mean): 27.0 Race: NR Gestational age at entry (mean): 291.4 days Dating criteria: LMP and either a) 1st trimester U/S or b) two 2nd trimester U/S Parity: 29% primiparous Bishop score: 41.5% > 6; 58.5% 3-6	1) Apgar score < 7 at 5 minutes 2) Macrosomia (> 4000 g) 3) C-sections 4) 2 x 2 tables 5) Other test performance results	1) Apgar score < 7 at 5 minutes: 1/65 (1.5%) 2) Macrosomia (> 4000 g): 10/65 (15.4%) 3) C-sections: Overall: 4/65 (6.2%) For fetal distress: 1/65 (1.5%) 4) 2 x 2 tables: <u>2 x 2 table 1:</u> Reference standard = C-section for fetal distress (not defined) Screening test = Fetal breathing movements (FBM) by U/S C-section yes / no / Totals: FBM - 0 / 24 / 24 FBM + 1 / 40 / 41 Totals: 1 / 64 / 65 <u>2 x 2 table 2:</u> Reference standard = Apgar score at 5 minutes Screening test = FBM Apgar at 5 min < 7 / ≥ 7 / Totals: FBM - 0 / 24 / 24 FBM + 1 / 40 / 41 Totals: 1 / 64 / 65 <u>2 x 2 table 3:</u> Reference standard = Macrosomia (birthweight > 4000 g) Screening test = FBM Macrosomia yes / no / Totals: FBM - 4 / 20 / 24 FBM + 6 / 35 / 41 Totals: 10 / 55 / 65	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: - *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Type(s) of providers: NR Length of follow-up: None			5) Other test performance results: Presence of fetal breath movements (FBM+) was associated with: a) No difference in birthweight (3608 ± 671 g vs. 3719 ± 710 g; p = not significant) b) Longer total induction time (648.5 ± 354 min vs. 319.3 ± 137 min; p < 0.001) c) Higher oxytocin requirement (2708 ± 1727 mIU vs. 1134 ± 709 mIU; p < 0.001)	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Sherer, Onyeije, Binder, et al., 1998	Design: Cohort/nested case-control study (retrospective) Test(s) studied: 1) FHR assessed for baseline fetal tachycardia (≥ 160 bpm) or bradycardia (≤ 120 bpm) Protocol: Baseline FHR assessed at post-term evaluation. Reference standard(s): 1) Apgar score at 5 minutes 2) Meconium aspiration 3) Admission to NICU Dates: July 1985 - June 1995 Location: Bronx, NY Setting: University hospital Type(s) of providers: MFM Length of follow-up: None	No. of subjects at start: 107 cases and 283 controls: 31 patients with baseline tachycardia, plus 66 matched controls; 76 patients with baseline brady-cardia, plus 217 matched controls Dropouts: NA (retrospective study) Loss to follow-up: NA No. of subjects at end: 107 cases and 283 controls Inclusion criteria: Singleton pregnancy; gestational age ≥ 41 weeks; not in labor; afebrile; normal fetal anatomy; reactive NST; intact membranes; no evidence of chorioamnionitis Exclusion criteria: Fetal tachy- or brady-arrhythmias; FHR decelerations; loss of short-term beat-to-beat variability Age: NR Race: NR Gestational age at entry: NR (gestational age ≥ 41 weeks required for entry into study) Dating criteria: LMP and U/S before 20 weeks Parity: NR Bishop score: NR	1) Apgar score < 7 at 5 minutes 2) Meconium staining 3) Meconium aspiration 4) Admission to NICU 5) Fetal growth restriction (< 10th percentile for gestational age) 6) C-sections 7) 2 x 2 tables	1) Apgar score < 7 at 5 minutes: Tachycardia: 7/31 (23%) Matched controls: 10/66 (15%) p = 0.369 Bradycardia: 14/76 (18%) Matched controls: 54/217 (25%) p = 0.25 2) Meconium staining: Tachycardia: 13/31 (42%) Matched controls: 28/66 (42%) p = 0.964 Bradycardia: 26/76 (34%) Matched controls: 63/217 (29%) p = 0.398 3) Meconium aspiration: Tachycardia: 2/31 (7%) Matched controls: 1/66 (2%) p = 0.190 Bradycardia: 4/76 (5%) Matched controls: 12/217 (6%) p = 0.929 4) Admission to NICU: Tachycardia: 2/31 (7%) Matched controls: 3/66 (5%) p = 0.692 Bradycardia: 11/76 (15%) Matched controls: 20/217 (9%) p = 0.199 5) Fetal growth restriction: Tachycardia: 1/31 (3%) Matched controls: 10/66 (15%) p = 0.084 Bradycardia: 8/76 (11%) Matched controls: 15/217 (7%) p = 0.313	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: - Dating criteria: + Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: - *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				6) C-sections: Tachycardia: 8/31 (25%) Matched controls: 11/66 (29%) p = 0.29 Bradycardia: 11/76 (15%) Matched controls: 52 217 (24%) p = 0.083 7) 2 x 2 tables: <u>2 x 2 table 1:</u> Reference standard = Apgar score at 5 minutes Screening test = Baseline bradycardia (BB) (≤ 120 bpm) Apgar at 5 min < 7 / ≥ 7 / Totals: BB yes 14 62 76 BB no 54 163 217 Totals: 68 225 293 <u>2 x 2 table 2:</u> Reference standard = Meconium aspiration Screening test = BB Meconium aspiration yes / no / Totals: BB yes 4 72 76 BB no 12 205 217 Totals: 16 277 293 <u>2 x 2 table 3:</u> Reference standard = Admission to NICU Screening test = BB NICU admission yes / no / Totals: BB yes 11 65 76 BB no 20 197 217 Totals: 31 262 293	*(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

2 x 2 table 4:
Reference standard = Apgar score at 5 minutes
Screening test = Baseline tachycardia (BT) (≥ 160 bpm)

	Apgar at 5 min		
	< 7	≥ 7	Totals:
BT yes	7	24	31
BT no	10	56	66
Totals:	17	80	97

2 x 2 table 5:
Reference standard = Meconium aspiration
Screening test = BT

	Meconium aspiration		
	yes	no	Totals:
BT yes	2	29	31
BT no	1	65	66
Totals:	3	94	97

2 x 2 table 6:
Reference standard = Admission to NICU
Screening test = BT

	NICU admission		
	yes	no	Totals:
BT yes	2	29	31
BT no	3	63	66
Totals:	5	92	97

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Small, Phelan, Smith, et al., 1987	Design: Case series (retrospective), historical controls Test(s) studied: 1) Nonstress test (n = 470) Protocol: For patients with good dates (U/S before 28 weeks or multiple 1st and 2nd trimester exams), NST performed twice weekly. If cervix favorable (Bishop score ≥ 9), then labor induced. For patients with unreliable dates (LMP only), NST performed weekly. NST considered reactive whenever ≥ 2 FHR accelerations observed within 10 minutes. Accelerations had to rise 15 bpm and last 15 seconds. Labor induced for FHR deceleration of any type; persistent nonreactive NST; oligohydramnios (< 1 cm) on U/S; positive contraction stress test (CST); or biophysical profile score ≤ 4. Reference standard(s): 1) C-section for fetal distress 2) Apgar score at 1 minute 3) Apgar score at 5 minutes 4) Macrosomia 5) Post-maturity Dates: Jan - Dec 1984 (study group); 1980 (controls) Location: Los Angeles, CA Setting: University hospital Type(s) of providers: General OB/GYN	No. of subjects at start: 476 cases (met inclusion criteria); 239 historical controls Dropouts: 6 cases (excluded due to incomplete delivery information) Loss to follow-up: NA No. of subjects at end: 470 cases; 239 historical controls Inclusion criteria: Gestational age > 294 days/42 weeks; antepartum FHR testing within 7 days of delivery Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry: NR (gestational age > 42 weeks required for entry into study) Dating criteria: LMP Parity: NR Bishop score: NR	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Meconium staining 4) Macrosomia (> 4000 g) 5) Post-maturity 6) Perinatal death 7) C-sections 8) 2 x 2 tables 9) Comparisons with historical controls	1) Apgar score < 7 at 1 minute: 86/470 (18%) 2) Apgar score < 7 at 5 minutes: 9/470 (2%) 3) Meconium staining: 126/470 (27%) 4) Macrosomia: 98/470 (21%) 5) Post-maturity: 32/470 (7%) 6) Perinatal death: 3/470 (< 1%) 7) C-sections: Overall: 79/470 (17%) For fetal distress: 19/470 (4%) 8) 2 x 2 tables: 2 x 2 table 1: Reference standard = C-section for fetal distress (not defined) Screening test = Nonstress test (NST) ("reactive" whenever ≥ 2 FHR accelera-tions observed within 10 minutes; accelerations had to rise 15 bpm and last 15 seconds) C-section for fetal distress yes / no / Totals: NST nonreactive 4 / 46 / 50 NST reactive 15 / 405 / 420 Totals: 19 / 451 / 470 2 x 2 table 2: Reference standard = Apgar score at 1 minute Screening test = NST (as above)	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: - Dating criteria: + Other risk factors absent: - Similar to likely pt pop: - Testing protocol described: - Sample size: - Statistical tests: - *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Length of follow-up: None			Apgar at 1 min < 7 ≥ 7 Totals: NST nonreactive 11 39 50 NST reactive 75 345 420 Totals: 86 384 470 2 x 2 table 3: Reference standard = Apgar score at 5 minutes Screening test = NST (as above) Apgar at 5 min < 7 ≥ 7 Totals: NST nonreactive 1 49 50 NST reactive 8 412 420 Totals: 9 461 470 2 x 2 table 4: Reference standard = Macrosomia (birthweight > 4000 g) Screening test = NST (as above) Macrosomia yes no Totals: NST nonreactive 5 45 50 NST reactive 93 327 420 Totals: 98 372 470 2 x 2 table 5: Reference standard = Post-maturity Screening test = NST (as above) Post-maturity yes no Totals: NST nonreactive 4 46 50 NST reactive 28 392 420 Totals: 32 438 470 9) Comparisons with historical controls: Compared to controls from 1980 (n = 239), post-dates patients from 1984	*(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				(n = 470) were significantly less likely to have meconium. Compared to controls, 1984 post-dates patients with reactive NST and decelerations were significantly less likely to have C-section for fetal distress, meconium, or birthweight > 4000 g. (Reactive NST with decelerations included among criteria for induction in 1984.)	

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Tam, Tai, and Rogers, 1999	Design: Cohort study (not specified if prospective) Test(s) studied: 1) Fetal fibronectin (fFN) testing, followed by induction using PGE_2 pessaries (n = 58) Protocol: Cervico-vaginal secretion tested for presence of fetal fibronectin prior to cervical ripening/induction. Labor induced with PGE_2 pessary (3 mg). Cervical status reassessed 4-6 hours later. If Bishop score < 5, then second dose given. If Bishop score ≥ 5, then artificial rupture of membranes performed. Oxytocin begun at 2.5 mU/min of 1 mU/min for nulliparous and multiparous women, respectively, with dose increased every 15 minutes. Reference standard(s): 1) C-section Dates: Apr 1996 - Feb 1997 Location: Hong Kong Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 58 (30 negative for fetal fibronectin [fFN-]; 28 positive [fFN+]) Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 58 Inclusion criteria: Term or post-term pregnancy; documented indication for induction Exclusion criteria: Bishop score ≥ 5; ruptured membranes Age (mean, with range): fFN-: 30 (27-33) fFN+: 28 (24-34) Race: NR Gestational age at entry (median, with range): fFN-: 281 days (272-294) fFN+: 294 days (280-294) p = 0.10 Dating criteria: NR Parity (median, with range): fFN-: 1 (0-1) fFN+: 1 (0-2) Bishop score (median, with range): fFN-: 3 (1-4) fFN+: 3 (1-4)	1) 2 x 2 table 2) Interval from induction to delivery	1) 2 x 2 table: Reference standard = C-section Screening test = Fetal fibronectin (fFN) status C-section yes / no / Totals: fFN+: 3 / 16 / 19 fFN-: 11 / 28 / 39 Totals: 14 / 44 / 58 2) Interval from induction to delivery (median, with range): fFN+: 760 minutes (540-1375) fFN-: 1285 minutes (692-2266) p = 0.04	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: + Dating criteria: - Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: + Results not stratified by indication for induction.

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Tongsong and Srisom-boon, 1993	Design: Cohort study (prospective) Test(s) studied: 1) Nonstress test (NST) + amniotic fluid volume (AFV) (n = 252) Protocol: Above-mentioned tests performed twice weekly. If NST or AFV abnormal, then contraction stress test (CST) performed. If CST negative, then patient re-tested in 3-4 days (uncertain if repeat test was NST+AFV or repeat CST). If CST positive, then labor induced. If cervix favorable, then labor induced. Reference standard(s): 1) Fetal distress/obstetric intervention 2) Apgar score at 1 minute 3) Apgar score at 5 minutes Dates: June 1989 - May 1992 Location: Chiang Mai, Thailand Setting: Unspecified hospital Type(s) of providers: MFM Length of follow-up: None	No. of subjects at start: 252 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 252 Inclusion criteria: Singleton pregnancy; attended antenatal clinic in 1st trimester; delivery after 42 weeks' gestation Exclusion criteria: Any medical or obstetric complication; congenital abnormalities of fetus Age: NR Race: NR Gestational age at entry: NR (delivery after 42 weeks required for entry into study) Dating criteria: LMP + 1st trimester clinical exam Parity: NR Bishop score: NR	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Fetal distress 4) Meconium staining 5) Obstetric intervention for fetal distress 6) 2 x 2 tables 7) Other test performance results	1) Apgar score < 7 at 1 minute: 17/252 (7%) 2) Apgar score < 7 at 5 minutes: 6/252 (2%) 3) Fetal distress: 11/252 (4%) 4) Meconium staining: 87/252 (35%) 5) Obstetric intervention for fetal distress: 11/252 (4%) 6) 2 x 2 tables: <u>2 x 2 table 1:</u> Reference standard = Obstetric intervention for fetal distress (defined as repetitive late decelerations, repetitive moderate to severe variable decelera-tions, or prolonged bradycardia) Screening test = Amniotic fluid volume (AFV) ("abnormal" if largest vertical pocket < 3 cm) Fetal distress yes / no / Totals: AFV abnormal: 8 / 22 / 30 AFV normal: 3 / 219 / 222 Totals: 11 / 241 / 252 <u>2 x 2 table 2:</u> Reference standard = Apgar score at 1 minute Screening test = AFV (as above) Apgar at 1 min < 7 / ≥ 7 / Totals: AFV abnormal: 8 / 22 / 30 AFV normal: 9 / 213 / 222 Totals: 17 / 235 / 252	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: + Dating criteria: + Other risk factors absent: + Similar to likely pt pop: - Testing protocol described: + Sample size: - Statistical tests: +

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

2 x 2 table 3:
Reference standard = Apgar score at 5 minutes
Screening test = AFV (as above)

	Apgar at 5 min		
	< 7	≥ 7	Totals:
AFV abnormal	2	28	30
AFV normal	4	218	222
Totals:	6	246	252

2 x 2 table 4:
Reference standard = Obstetric intervention for fetal distress (as above)
Screening test = Nonstress test (NST) ("abnormal" if nonreactive or reactive with variable or late decelerations)

	Fetal distress		
	yes	no	Totals:
NST abnormal	7	44	51
NST normal	4	197	201
Totals:	11	241	252

2 x 2 table 5:
Reference standard = Apgar score at 1 minute
Screening test = NST (as above)

	Apgar at 1 min		
	< 7	≥ 7	Totals:
NST abnormal	7	44	51
NST normal	10	191	201
Totals:	17	235	252

2 x 2 table 6:
Reference standard = Apgar score at 5 minutes
Screening test = NST (as above)

(continued on next page)

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

	Apgar at 5 min		
	< 7	≥ 7	Totals:
NST abnormal	2	49	51
NST normal	4	197	201
Totals:	6	246	252

7) Other test performance results:
AFV was more accurate than NST in predicting intrapartum fetal distress ($p < 0.05$).

AFV sensitivity, 0.73; specificity, 0.91; positive predictive value, 0.27; negative predictive value, 0.99.

NST sensitivity, 0.64; specificity, 0.82; positive predictive value, 0.14; negative predictive value, 0.98.

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Weiner, Farmakides, Schulman, et al., 1994	Design: Cohort study (prospective) Test(s) studied: 1) Nonstress test (NST) with computerized analysis of fetal heart rate (FHR) variation + Doppler examination of umbilical artery + biophysical profile (n = 337) Protocol: Above-mentioned tests performed every 2-4 days beginning at 41 weeks. Labor induced (using oxytocin infusion and amniotomy) if FHR variation reduced (< 30 msec), FHR decelerations appeared, or amniotic fluid index (AFI) ≤ 5, and after 42 weeks if Bishop score > 7. Reference standard(s): 1) Fetal distress 2) Acidosis 3) Neonatal death Dates: June 1991 - May 1993 Location: Mineola, NY Setting: University hospital Type(s) of providers: MFM Length of follow-up: None	No. of subjects at start: 337 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 337 Inclusion criteria: Delivery at > 41 weeks' gestation; uncomplicated pregnancy Exclusion criteria: None specified Age (mean ± SD): 29 ± 4.6 Race: NR Gestational age at entry: NR (delivery at > 41 weeks required for entry into study) Dating criteria: U/S before 22 weeks Parity: NR Bishop score: NR	1) Fetal distress 2) Acidosis 3) Neonatal death 4) C-sections 5) 2 x 2 tables 6) Other test performance results	1) Fetal distress: 37/337 (11%) 2) Acidosis: 10/337 (3%) 3) Neonatal death: 2/337 (0.6%) 4) C-sections: Overall: 101/337 (30%) For fetal distress: 33/337 (10%) 5) 2 x 2 tables: <u>2 x 2 table 1:</u> Reference standard = Fetal distress (definition included presence of FHR late decelerations, severe FHR variable decelerations, and reduced beat-to-beat variability) Screening test = FHR variation Fetal distress yes / no / Totals: FHR variation < 30 msec: 11 / 1 / 12 FHR variation ≥ 30 msec: 28 / 297 / 325 Totals: 39 / 298 / 337 <u>2 x 2 table 2:</u> Reference standard = Fetal acidosis (umbilical artery pH < 7.2) Screening test = FHR variation Acidosis yes / no / Totals: FHR variation < 30 msec: 7 / 5 / 12 FHR variation ≥ 30 msec: 3 / 322 / 325 Totals: 10 / 327 / 337 <u>2 x 2 table 3:</u> Reference standard = Neonatal death Screening test = FHR variation *(continued on next page)*	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: + Gestational age: + Dating criteria: + Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: + Statistical tests: +

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

	Neonatal death		
	yes	no	Totals:
FHR variation < 30 msec	1	11	12
FHR variation ≥ 30 msec	1	324	325
Totals:	2	335	337

6) Other test performance results:
For predicting intrapartum fetal distress and acidosis at delivery, FHR variations showed higher area under the ROC curve than did amniotic fluid index or umbilical S:D ratio. Nonreactive NST and presence of decelerations were also predictive of distress in labor, and decelerations were predictive of acidosis at delivery ($p < 0.001$).

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Weiner, Reichler, Zlozover, et al., 1993	Design: Cohort study (prospective) Test(s) studied: 1) Nonstress test (NST) + amniotic fluid volume (AFV) + Doppler velocimetry of umbilical and uterine arteries (n = 142) Protocol: Above-mentioned tests performed every 3 days. Labor induced if abnormal NST, oligohydramnios, or favorable cervix (Bishop score > 7) after 42 weeks gestation. Reference standard(s): 1) Fetal outcome Dates: NR; data collected over a 1-year period Location: Haifa, Israel Setting: University hospital Type(s) of providers: General OB/GYN Length of follow-up: None	No. of subjects at start: 142 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 142 Inclusion criteria: Gestational age > 287 days Exclusion criteria: Pregnancy complications (e.g., hypertension, gestational diabetes) Age (mean ± SD): 27.3 ± 5.6 Race: NR Gestational age at entry: NR (gestational age > 287 days required for entry into study) Dating criteria: "Early fetal biometry" Parity: 31% primiparous Bishop score: NR	1) Apgar score < 7 at 5 minutes 2) Admission to NICU 3) C-sections 4) 2 x 2 tables 5) Other test performance results	1) Apgar score < 7 at 5 minutes: 2/142 (1.4%) 2) Admission to NICU: 1/142 (0.7%) 3) C-sections: Overall: 13/142 (9.2%) For fetal distress: 7/142 (4.9%) 4) 2 x 2 tables: <u>2 x 2 table 1:</u> Reference standard = Fetal outcome ("abnormal" defined as 5-minute Apgar score < 7, admission to NICU, C-section for fetal distress, or birthweight < 5th percentile) Screening test = NST Fetal outcome abn nl Totals: NST abnormal 1 6 7 NST normal 11 124 135 Totals: 12 130 142 <u>2 x 2 table 2:</u> Reference standard = Fetal outcome (as above) Screening test = AFV ("low" defined as < 5 cm) Fetal outcome abn nl Totals: AFV low 3 8 11 AFV normal 9 122 131 Totals: 12 130 142 <u>2 x 2 table 3:</u> Reference standard = Fetal outcome (as above) Screening test = NST, AFV, and umbilical and uterine artery resistance index	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: + Gestational age: + Dating criteria: + Other risk factors absent: - Similar to likely pt pop: + Testing protocol described: + Sample size: - Statistical tests: - *(continued on next page)*

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes

	Fetal outcome abn	nl	Totals:
At least one test abn	8	18	26
All tests normal	4	112	116
Totals:	12	130	142

5) Other test performance results:
For predicting abnormal fetal outcome (as defined above), screening tests had the following performance characteristics:

	Sensitivity	Specificity
NST	0.08	0.95
AFV	0.25	0.94
Resistance index	0.17	0.96
Any test abnormal	0.67	0.88

Evidence Table 1: Studies relevant to Key Question 1 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Witter and Weitz, 1989	Design: Case series (prospective), no controls Test(s) studied: 1) Cervical exam + induction by oxytocin infusion and amniotomy (n = 76) Protocol: At 42 completed weeks, cervical exam performed prior to induction of labor. Oxytocin infusion started at 7:00 AM with 1 mU/min and increased by 1 mU/min every 10 min until a dose of 30 mU/min reached or a regular pattern of adequate uterine contractions established. Amniotomy performed as soon as possible, but always after oxytocin had established regular contractions. If patient had intact membranes and was not in active phase labor by evening, the induction was stopped and the patient was rested overnight. The induction was restarted in the morning. If the patient failed to enter the active phase of labor by 20 hours of induction, then C-section performed. Reference standard(s): 1) C-section Dates: NR Location: Baltimore, MD Setting: University hospital Type(s) of providers: MFM Length of follow-up: None	No. of subjects at start: 103 (see Notes) Dropouts: 27 (did not have cervical exam) Loss to follow-up: NA No. of subjects at end: 76 Inclusion criteria: Gestational age ≥ 42 weeks Exclusion criteria: Previous C-section Age: NR Race: NR Gestational age at entry: NR (gestational age ≥ 42 weeks required for entry into study) Dating criteria: 2 or more of the following: certain LMP; basal body temperature indicating ovulation temperature shift for the present pregnancy; positive urinary pregnancy test at 6 weeks from LMP; fetal heart tones heard with DeLee stethoscope at 18-20 weeks; fundal height at the umbilicus at 20 weeks; fundal height in cm equal to gestational age in weeks within 2 cm from 20-34 weeks; early registration with dates equal to exam prior to 13 weeks; U/S dating by crown-rump length between 6 and 14 weeks or by biparietal diameter prior to 26 weeks Parity: NR Bishop score: NR	1) C-sections 2) 2 x 2 tables	1) C-sections: 26/76 (34%) 2) 2 x 2 tables: 2 x 2 table 1: Reference standard = C-section Screening test = Cervical dilation C-section yes / no / Totals: Dilation 0 cm: 20 / 11 / 31 Dilation > 0 cm: 6 / 39 / 45 Totals: 26 / 50 / 76 2 x 2 table 2: Reference standard = C-section Screening test = Cervical effacement C-section yes / no / Totals: Effacement 0%: 12 / 6 / 18 Effacement > 0%: 14 / 44 / 58 Totals: 26 / 50 / 76 2 x 2 table 3: Reference standard = C-section Screening test = Cervical station C-section yes / no / Totals: Station ≥ -3: 19 / 22 / 41 Station < -3: 7 / 28 / 35 Totals: 26 / 50 / 76	QUALITY SCORE: Reference standard: + Randomized: - Method of randomization: NA Verification bias: - Test reliability/variability: - Gestational age: + Dating criteria: + Other risk factors absent: - Similar to likely pt pop: - Testing protocol described: + Sample size: - Statistical tests: + Study population was subgroup (76/103) of patients randomized to induction in Witter and Weitz, 1987 (see Evidence Table: Key Question 3).

Evidence Table 2: Studies relevant to Key Question 2

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Augensen, Bergsjø, Eikeland, et al., 1987	Design: RCT, randomization by random numbers list Interventions: 1) Immediate induction at time of referral/admission into study (n = 214) Protocol: 5 IU oxytocin given intravenously, with dose rates increased stepwise according to response. Amniotomy performed once labor established or, in exceptional cases, at the start of induction. If no labor after 6-8 hours, then induction considered unsuccessful, and patient managed according to postponed induction protocol. 2) Delayed induction after monitoring for 1 wk (n = 195) Protocol: NST on day of referral/admission into study and again on day 3 or 4 if still undelivered. If birth had not occurred by day 7, then labor induced as above. If this induction attempt failed, then management "left to clinical judgement." Dates: Jan 1982 - June 1985 Location: Bergen, Norway Setting: University hospital Type(s) of providers: NR Length of follow-up: NA	No. of subjects at start: 409 Dropouts: 0 (see notes) Loss to follow-up: NA No. of subjects at end: 409 Inclusion criteria: Healthy women with normal pregnancies; singleton fetus; cephalic presentation; gestational age 290-297 days; reliable dates Exclusion criteria: Use of OCPs during two months before LMP; hypertension; IUGR; other medical conditions; geographical and social considerations (not specified) Age: NR Race: NR Gestational age at entry: NS; gestational age of 290-297 days required for entry into study Dating criteria: LMP ("clear recollection") Parity: Immediate induction, 46% nulliparous; delayed induction, 42% Bishop score: Immediate induction, 36% < 6; delayed induction, 35% < 6	1) Meconium 2) Admission to NICU 3) Length of stay in NICU 4) Hyperbilirubinemia 5) Difficult shoulder delivery 6) C-sections 7) Number of days in hospital 8) Courses of induction	1) Meconium: Immediate: 37/214 (17%) Delayed: 32/195 (16%) (no p-value reported) 2) Admission to NICU: Immediate: 12/214 (5.6%) Delayed: 15/195 (7.7%) (no p-value reported) 3) Length of stay in NICU (mean): Immediate: 4.3 days Delayed: 9.7 days (one patient stayed in NICU 93 days) (no p-value reported) 4) Hyperbilirubinemia: Immediate: 10/214 (4.7%) Delayed: 1/195 (0.51%) $0.01 > p > 0.005$ 5) Difficult shoulder delivery: Immediate: 1/214 (0.5%) Delayed: 0/195 (no p-value reported) 6) C-sections: Immediate: 14/214 (6.5%) Delayed: 15/195 (7.7%) (no p-value reported) 7) Number of days in hospital (mean $\pm$ SD): Immediate: 7.05 ± 1.67 days Delayed: 6.69 ± 1.37 days $p = 0.02$ 8) Courses of induction (mean): Immediate: 1.09 Delayed: 0.34 (no p-value reported)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: - Four patients randomized to immediate induction delivered spontaneously before being induced; these patients were included in the analysis. Results not stratified by parity.

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Bergsø, Huang, Yu, et al., 1989	Design: RCT, randomization by list of random numbers Interventions: 1) Induction (n = 94) Protocol: Labor induced at or shortly after 42 weeks by stripping of the membranes, followed by oxytocin infusion (5 IU in 500 ml solution). Infusion rate regulated according to response. Membranes ruptured artificially if cervix dilated ≥ 3 cm. 2) Monitoring (n = 94) Protocol: Patients admitted to hospital to undergo "close daily clinical surveillance." Fetal movement tests, atropine tests, U/S, and urinary estriol excretion tests also employed. Labor induced as above at ≥ 43 weeks "according to clinical judgement." Dates: July 1982 - sometime in 1984 Location: Wuhan, China Setting: Community hospital which also serves as regional referral center for high-risk obstetrics Type(s) of providers: Unspecified OB/GYN Length of follow-up: NA	No. of subjects at start: 188 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 188 Inclusion criteria: Gestational age ≥ 42 weeks (294 days); not in labor; intact membranes; normal pregnancy without significant risk factors; normal menstrual cycle (28 ± 4 days) with accurate recall of LMP Exclusion criteria: None specified Age (mean): Induction, 26.1; monitoring, 27.8 Race: 100% Chinese Gestational age at entry: NR; gestational age of ≥ 42 weeks required for entry into study Dating criteria: LMP Parity: Induction, 6% nulliparous; monitoring, 13% nulliparous Bishop score: NR	1) Apgar scores 2) Fetal distress 3) Hyperbilirubinemia 4) Respiratory distress syndrome 5) Aspiration pneumonia 6) Total operative deliveries (C-sections, forceps-assisted deliveries, and vacuum extractions) 7) C-sections 8) Forceps-assisted deliveries 9) Vacuum extractions 10) Length of hospital stay	1) Apgar scores: No quantitative data reported. Authors stated only that "Apgar score distributions were almost equal between the groups." 2) Fetal distress (not defined): Induction: 17/94 (18.1%) Monitoring: 18/94 (19.1%) p = not significant 3) Hyperbilirubinemia: Induction: 6/94 (6.4%) Monitoring: 3/94 (3.2%) p = not significant 4) Respiratory distress syndrome: Induction: 4/94 (4.3%) Monitoring: 8/94 (8.5%) p = not significant 5) Aspiration pneumonia: Induction: 4/94 (4.3%) Monitoring: 8/94 (8.5%) p = not significant 6) Total operative deliveries: Induction: 48/94 (51.1%) Monitoring: 64/94 (68.1%) p < 0.05 7) C-sections: Induction: 27/94 (28.7%) Monitoring: 39/94 (41.5 %) p = not significant 8) Forceps-assisted deliveries: Induction: 9/94 (9.6%) Monitoring: 11/94 (11.7%) p = not significant 9) Vacuum extractions: Induction: 12/94 (12.8%) Monitoring: 14/94 (14.9%)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: - Mode of delivery: - Sample size: - Statistical tests: ?? Gestational age: + Dating criteria: + Bishop score: - Results not stratified by parity. *(continued on next page)*

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				p = not significant 10) Length of hospital stay (mean, with range) Induction: 7.9 days (1-28) Monitoring: 8.1 days (1-22 (no p-value reported)	

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Cardozo, Fysh, and Pearce, 1986 **(Original intention-to-treat analysis)** ***and*** **Pearce and Cardozo, 1988** **(Supplementary analysis including only patients who actually received the treatment to which they were allocated)**	Design: RCT, randomization by chart number Interventions: 1) Planned induction (n = 195 [intention-to-treat analysis]; 125 [supplemental analysis]) Protocol: Labor induced between 40 weeks + 12 days and 40 weeks + 14 days (2-4 days after recruitment/ randomization). PGE_2 suppository (3 mg) inserted, followed 3 hours later by amniotomy and, if necessary, oxytocin infusion. 2) Expectant management (n = 207 [intention-to-treat analysis]; 156 [supplemental analysis]) Protocol: U/S exam given between 40 weeks + 12 days and 40 weeks + 16 days (2-6 days after recruitment/ randomization) to determine ratio of head circumference to abdominal circumference and to estimate amniotic fluid volume. Patients monitored with daily kick count charts and cardiotocography on alternate days. Labor induced for asymmetric IUGR with abnormal cardiotocogram; PROM; or onset of hypertension. Patients in both groups were permitted to request or decline induction of labor after 42 weeks' gestation.	No. of subjects at start: 402 Dropouts: 70 patients in the active group and 41 or 51 in the expectant management group. According to the original publication (Cardozo, Fysh, and Pearch, 1986), 49/70 dropouts from the active group went into labor spontaneously during the waiting period before the planned induction, while the other 21 asked to be induced. According to the supplementary analysis (Pearce and Cardozo, 1988), all 70 went into labor spontaneously. According to the original publication, 2/41 dropouts in the expectant management group had elective C-sections, while the remaining 39 were induced during the waiting period. According to the supplementary analysis, 41 women in the expectant management group went into spontaneous labor during the waiting period, and an additional 10 were induced during the waiting period. All these patients were included in the original intention-to-treat analysis, but were excluded from the later supplementary analysis. Demographic data below are for the intention-to-treat population. Loss to follow-up: NA No. of subjects at end: 402 (intention-to-treat analysis); 281 (supplemental analysis) Inclusion criteria: Uncomplicated pregnancy; gestational age 40	1) Apgar score < 5 at 1 minute 2) Apgar score < 5 at 5 minutes 3) Birthweight 4) Cord venous pH 5) Meconium aspiration syndrome 6) Major FHR tracing abnormality 7) Admission to NICU 6) Duration of 2nd stage of labor 7) Intervention during 2nd stage of labor 8) Forceps-assisted delivery 9) Emergency C-sections 10) Patient satisfaction	1) Apgar score < 5 at 1 minute: *Intention-to-treat analysis:* Induction: 30/195 (15%) Expectant mgmt: 25/207 (12%) p = not significant *Supplemental analysis:* Induction: 19/125 (15%) Expectant mgmt: 16/156 (10%) p = not significant 2) Apgar score < 5 at 5 minutes: *Intention-to-treat analysis:* Induction: 2/195 (1%) Expectant mgmt: 4/207 (2%) p = not significant *Supplemental analysis:* Induction: 1/125 (1%) Expectant mgmt: 2/156 (1%) p = not significant 3) Birthweight (mean ± SD): *Intention-to-treat analysis:* Induction: 3.69 ± 0.51 kg Expectant mgmt: 3.63 ± 0.43 kg p = not significant *Supplemental analysis:* Induction: 3670 ± 500 g Expectant mgmt: 3630 ± 400 g p = not significant 4) Cord venous pH (mean ± SD): *Intention-to-treat analysis:* Induction (n = 84): 7.29 ± 0.10 Expectant mgmt (n = 99): 7.32 ± 0.08 p < 0.05 *Supplemental analysis:* Induction: 7.28 ± 0.10 Expectant mgmt: 7.33 ± 0.08 p = 0.006	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: - Differences exist between the original and supplementary articles in reporting of the number of patients who went into spontaneous labor before the planned induction period. Original article: 49 (induction group) vs. 0 (expectant management group). Supplementary article: 70 (induction group) vs. 41 (expectant management group) ($p < 0.05$). Significant difference between two groups in racial distribution at baseline. Results not stratified by parity. No data on baseline Bishop scores. *(continued on next page)*

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Dates: NR (patients enrolled over a 21-month period) Location: London, England Setting: 2 hospitals of unspecified type Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	weeks + 10 days (290 days) Exclusion criteria: None specified Age: NR; authors stated only that two groups were "well matched" for maternal age Race: Induction, 73% White; expectant management, 83% White ($p < 0.05$) Gestational age at entry (mean ± SD): 290 days (inclusion criterion) Dating criteria: LMP and U/S performed before 20 weeks Parity: NR; authors stated only that two groups were "well matched" for parity Bishop score: Baseline scores NR		5) Meconium aspiration syndrome: *Intention-to-treat analysis:* Induction: 1/195 (0.5%) Expectant mgmt: 1/207 (0.5%) p = not significant *Supplemental analysis:* Induction: 4/125 (3%) Expectant mgmt: 5/156 (1%) p = not significant 6) Major FHR tracing abnormality: *Intention-to-treat analysis:* Induction: 27/195 (14%) Expectant mgmt: 17/207 (8%) p = not significant *Supplemental analysis:* Induction: 22/125 (14%) Expectant mgmt: 11/156 (7%) $p < 0.02$ 7) Admission to NICU: *Intention-to-treat analysis:* Induction: 6/195 (3%) Expectant mgmt: 3/207 (1.5%) p = not significant *Supplemental analysis:* Induction: 5/125 (4%) Expectant mgmt: 1/156 (1%) p = not significant 6) Duration of 2nd stage of labor (mean): *Intention-to-treat analysis:* Induction (n = 175): 72 minutes Expectant mgmt (n = 188): 77 minutes p = not significant *Supplemental analysis:* Induction (n = 108): 66.2 minutes Expectant mgmt (n = 141): 78.8 minutes p = not significant	*(continued on next page)*

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				7) Intervention during 2nd stage of labor: *Intention-to-treat analysis:* Induction (n = 175): 44/175 (25%) Expectant mgmt (n = 188): 54/188 (29%) p = not significant *Supplemental analysis:* Induction (n = 108): 31/108 (29%) Expectant mgmt (n = 141): 40/141 (28%) p = not significant 8) Forceps-assisted delivery: *Intention-to-treat analysis:* Induction: 39/195 (20%) Expectant mgmt: 54/207 (26%) p = not significant *Supplemental analysis:* Induction (n = 108): 28/108 (26%) Expectant mgmt (n = 141): 39/141 (28%) p = not significant 9) Emergency C-sections: *Intention-to-treat analysis:* Induction: 25/195 (13%) Expectant mgmt: 18/207 (9%) p = not significant *Supplemental analysis:* Induction (n = 108): 3/108 (3%) Expectant mgmt (n = 141): 1/141 (1%) p = not significant 10) Patient satisfaction: *Intention-to-treat analysis:* Induction / ExpMgmt Pleased 49% / 53% No comment 34% / 35% Disappointed 15% / 11% No response 3% / 1% p = not significant *Supplemental analysis:* Not reported	

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Dyson, Miller, and Armstrong, 1987	Design: RCT, randomization according to table of random numbers and sealed envelopes Interventions: 1) Cervical ripening and induction (n = 152) Protocol: Patients underwent cervical ripening with PGE_2 gel (3 mg in initial phase of study, later changed to 0.5 mg), applied intravaginally on an outpatient basis. Patients monitored for ≥ 45 minutes. Those with regular contractions admitted to hospital for continued observation; others allowed to go home. If no labor the next morning (16-18 hours later), then patient admitted to hospital. Oxytocin induction begun if cervical score ≥ 5. If cervical score < 5, then second dose of PGE_2 gel administered and patient monitored for 4 hours. After 4 hours, oxytocin induction started regardless of cervical score. 2) Antepartum monitoring (n = 150) Protocol: NST performed twice weekly. Pelvic exam and determination of AFV performed weekly between 41 and 42 weeks gestation and twice weekly after 42 weeks. Labor induced if abnormal results on fetal testing or if cervical score became ≥ 6.	No. of subjects at start: 302 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 302 Inclusion criteria: Gestational age ≥ 287 days; low risk; unfavorable cervix Exclusion criteria: Risk factors known to increase perinatal mortality and morbidity (e.g., chronic hypertension, pre-eclampsia, diabetes, growth retardation, previous stillbirth); risk factors known to increase risk of induction (e.g., multiple gestation and polyhydramnios); risk factors know to affect C-section rate (e.g., breech presentation and previous C-section); favorable cervix (cervical score ≥ 6); nonreactive NST; variable deceleration on NST; oligohydramnios Age (mean ± SD): Induction, 24.8 ± 4.8; monitoring, 25.1 ± 5.0 Race: NR Gestational age at entry (mean ± SD): Induction, 290.8 ± 2.8 days; monitoring, 290.5 ± 2.6 days Dating criteria: 1) LMP confirmed by either a positive urine test within ≤ 6 weeks gestation or a 1st trimester pelvic exam or a 1st or 2nd trimester U/S; *or* 2) serial U/S exams, with the first performed before 24 weeks	1) Perinatal death 2) Apgar score < 7 at 1 minute 3) Apgar score < 7 at 5 minutes 4) Meconium staining 5) Meconium aspiration (meconium below the vocal cords on intubation, with admission to the NICU for oxygen administration) 6) Post-maturity syndrome 7) Fetal distress (abnormality of FHR tracing prompting C-section or midforceps delivery) 8) Birthweight 9) Macrosomia 10) C-sections 11) Maternal hospital stay 12) Infant hospital stay	1) Perinatal death: Induction: 0 Monitoring: 1/150 (< 1%) p = not significant 2) Apgar score < 7 at 1 minute: Induction: 17/152 (11.2%) Monitoring: 32/150 (21.3%) p < 0.02 3) Apgar score < 7 at 5 minutes: Induction: 2/152 (1.3%) Monitoring: 3/150 (2%) p = not significant 4) Meconium staining: Induction: 29/152 (19.1%) Monitoring: 70/150 (46.7%) p < 0.01 5) Meconium aspiration: Induction: 0 Monitoring: 6/150 (4.0%) p < 0.02 6) Post-maturity syndrome: Induction: 8/152 (5.3%) Monitoring: 22/150 (14.7%) p < 0.01 7) Fetal distress: Induction: 4/152 (2.6%) Monitoring: 27/150 (18.0%) p < 0.01 8) Birthweight (mean ± SD): Induction: 3696 ± 370 g Monitoring: 3766 + 428 p = not significant 9) Macrosomia: Induction: 29/152 (19.1%) Monitoring: 42/150 (28.2%) p = not significant	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + *(continued on next page)*

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Dates: Jan 1983 - Dec 1985 Location: Santa Clara, CA Setting: Community hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	Parity (mean ± SD): Induction, 0.4 ± 0.7 (70% nulliparous); monitoring, 0.3 ± 0.6 (73% nulliparous) Bishop score: NR (though see inclusion and exclusion criteria)		10) C-sections: *Overall:* Induction: 22/152 (14.5%) Monitoring: 41/150 (27.3%) p < 0.01 *Among nulliparous women:* Induction: 21/106 (19.8%) Monitoring: 38/110 (34.6%) p < 0.02 *Among multiparous women:* Induction: 1/46 (2.2%) Monitoring: 3/40 (7.5%) p = not significant 11) Maternal hospital stay (mean ± SD): Induction: 3.2 ± 1.3 days Monitoring: 3.5 ± 1.2 days p < 0.04 12) Infant hospital stay (mean ± SD): Induction: 3.0 ± 1.2 days Monitoring: 3.3 ± 1.5 days p = not significant	

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Egarter, Kofler, Fitz, et al., 1989	Design: RCT, method of randomization not described Interventions: 1) Induction of labor at due date by intravaginal PGE_2 tablets (3 mg) (n = 180) Protocol: 3 mg PGE_2 tablets applied vaginally. Dose repeated at 6 hours if labor did not start or contractions were inadequate. If patient still undelivered at 24 hours, but cervix ≥ 3 cm, then another treatment course given. If cervix < 3 cm, no further induction attempt performed. 2) "Watchful waiting" (n = 165) Protocol: Cardiotocographic evaluation of fetal well-being performed at 2- to 3-day intervals. Labor induced as above at completion of 42 weeks of amenorrhea. Dates: NR Location: Vienna, Austria Setting: University hospital Type(s) of providers: NR Length of follow-up: None	No. of subjects at start: 356 Dropouts: 11 Loss to follow-up: NA No. of subjects at end: 345 Inclusion criteria: Singleton pregnancies in cephalic presentation reaching their estimated date of confinement; intact membranes; cervix favorable for induction (modified Bishop score > 4) Exclusion criteria: Any fetal or maternal risk factor Age: NR Race: NR Gestational age at entry: NR Dating criteria: "Early" U/S Parity: Induction, 55% nulliparous; watchful waiting, 53% nulliparous Bishop score: NR	1) Fetal death 2) Other fetal outcomes 3) C-sections 4) Forceps-assisted delivery 5) Time from initial visit to spontaneous onset of labor (watchful waiting group only) 6) Number of pregnancies undelivered at 294 days in watchful waiting group 7) Use of analgesic treatment during labor	No p-values reported for outcomes described below. 1) Fetal death: Induction: 0 Watchful waiting: 1/165 (< 1%) 2) Other fetal outcomes: No significant differences between the two groups for birthweight and length, meconium staining, low Apgar scores, or pH. No quantitative data reported for these outcomes. 3) C-sections: *Among primiparae:* Induction: 1/99 (1.0%) Watchful waiting: 3/88 (3.4%) *Among multiparae:* Induction: 1/81 (1.2%) Watchful waiting: 0/77 4) Forceps-assisted delivery: *Among primiparae:* Induction: 3/99 (3.0%) Watchful waiting: 3/88 (3.4%) *Among multiparae:* Induction: 1/81 (1.2%) Watchful waiting: 0/77 5) Time from initial visit to spontaneous onset of labor (mean ± SD) (watchful waiting group only): Among nulliparae (n = 81): 4.5 ± 3.7 days Among multiparae (n = 75): 3.9 ± 2 days 6) Number of pregnancies undelivered at 294 days in watchful waiting group: 7/165 pregnancies (4.2%). All 7 deliveries were "uneventful," though	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: - Dating criteria: + Bishop score: + 11 patients crossed over, but were dropped from analysis. *(continued on next page)*

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				umbilical artery pH was slightly low (7.23) in one case. 7) Use of analgesic treatment during labor: Induction: 35% Watchful waiting: 35%	

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
El-Torkey and Grant, 1992	Design: RCT, randomization by random permuted blocks and sealed envelopes Interventions: 1) Sweeping of the membranes (n = 33) Protocol: Examination gloves lubricated with jelly or obstetric cream. As much of the membranes as possible were separated from the lower segment. If the cervix would not admit a finger, it was massaged vigorously to encourage prostaglandin release. Patients given date for formal induction of labor. 2) Monitoring (n = 32) Protocol: No form of vaginal examination given. No further details provided on management protocol. Patients given date for formal induction of labor. Dates: June 1990 - Mar 1991 Location: Bellshill, UK Setting: Community hospital Type(s) of providers: NR Length of follow-up: None	No. of subjects at start: 65 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 65 Inclusion criteria: Between 41 and 42 weeks gestation; preferred induction to monitoring when given choice Exclusion criteria: None specified Age (mean ± SD): Sweeping, 27.2 ± 4.7; monitoring, 25.3 ± 5.1 Race: NR Gestational age at entry (mean ± SD): Sweeping, 286.6 ± 2.8 days; monitoring, 286.3 ± 2.8 Dating criteria: NR Parity: Sweeping, 52% nulliparous; monitoring, 44% nulliparous Bishop score: NR	1) Apgar score < 6 at 1 minute 2) Apgar score < 6 at 5 minutes 3) Serious infection 4) Perinatal death 5) Maternal fever (axillary temperature > 37.1° C) 6) C-sections 7) Forceps-assisted delivery 8) Spontaneous delivery 9) Spontaneous labor	1) Apgar score < 6 at 1 minute: Sweeping: 2/33 (6%) Monitoring: 6/32 (19%) p = 0.12 2) Apgar score < 6 at 5 minutes: Sweeping: 1/33 (3%) Monitoring: 1/32 (3%) p = 0.98 3) Serious infection: Sweeping: 0 Monitoring: 0 4) Perinatal death: Sweeping: 0 Monitoring: 0 5) Maternal fever: Sweeping: 0 Monitoring: 4/32 (12.5%) p = 0.04 6) C-sections: Sweeping: 5/33 (15%) Monitoring: 4/32 (12.5%) p = 0.76 7) Forceps-assisted delivery: Sweeping: 2/33 (6%) Monitoring: 3/32 (9%) p =0.62 8) Spontaneous delivery: Sweeping: 26/33 (79%) Monitoring: 25/32 (78%) p =0.95 9) Spontaneous labor: Sweeping: 25/33 (76%) Monitoring: 12/32 (38%) p =0.002	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: - Patients in control group not informed that they were taking part in a randomized trial. Trial suspended before reaching n = 110 because of discrepancy in spontaneous labor rates (main outcome). No specific mention of use of Bishop score, except in reference to other studies that did not use it. Sample size estimates based on proportion of patients entering spontaneous labor.

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Gonen, Rosen, Dolfin, et al., 1997	Design: RCT, randomization by randomly generated numbers; method of concealment NR Interventions: 1) Induction of labor using oxytocin or prostaglandins, depending on Bishop score (criteria not specified) (n = 140) 2) Expectant management with NST/biophysical profile twice weekly and induction if no labor by 42 weeks (n = 144) Dates: Feb 1992 - Aug 1995 Location: Kfar-Saba, Israel Setting: Community hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 284 Dropouts: 11 Loss to follow-up: NA No. of subjects at end: 273 Inclusion criteria: Referral for ultrasound evaluation for potential macrosomia; completed 38 weeks; ultrasound EFW between 4,000 and 4,500 grams Exclusion criteria: Active labor; diabetes; prior cesarean delivery; nonvertex presentation; indications for induction other than macrosomia Age (mean ± SD): Induction, 30.8 ± 5.0; expectant, 29.5 ± 5.2 (p = 0.02) Race: NR Gestational age at entry (mean ± SD): Induction, 284.1 ± 6.4 days; expectant, 284.4 ± 5.7 days Dating criteria: NR Parity: Induction, 31% nulliparous; expectant, 29% nulliparous Bishop score: NR Other: Among nulliparous women, expectantly managed women younger (24.7 ± 3.0 vs. 27.6 ± 4.6; p = 0.001); no other differences	1) Time to delivery 2) Vaginal deliveries, stratified by parity 3) Instrumental deliveries, stratified by parity 4) C-section rates, stratified by parity 5) Umbilical artery pH 6) Shoulder dystocia 7) Cephalohematoma 8) Clavicular fracture 9) Brachial plexus palsy 10) Intraventricular hemorrhage	1) Time to delivery: Induction: 18.6 hours; range, 2-72 hours; 78% delivered within 24 hours Expectant: 4.1 ± 4.0 days Results similar in nulliparous and parous women 2) Vaginal deliveries, stratified by parity: *Overall:* Induction: 67.9% Expectant: 65.5% p = not significant *Nulliparous:* Induction: 35.7% Expectant: 50.0% p = not significant *Multiparous:* Induction: 82.6% Expectant: 71.7% p = not significant 3) Instrumental deliveries, stratified by parity: *Overall:* Induction: 12.7% Expectant: 12.9% p = not significant *Nulliparous:* Induction: 26.2% Expectant: 15.0% p = not significant *Multiparous:* Induction: 6.5% Expectant: 12.1% p = not significant 4) C-section rates, stratified by parity: *Overall:* Induction: 19.4% Expectant: 21.6%	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: - Study underpowered to detect differences in categorical variables and rare outcomes. Unclear if any women randomized to expectant management who were induced because of abnormal testing were excluded from analysis. *(continued on next page)*

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				p = not significant *Nulliparous:* Induction: 38.1% Expectant: 35.0% p = not significant *Multiparous:* Induction: 10.9% Expectant: 16.2% p = not significant 5) Umbilical artery pH: Induction: 7.32 ± 0.07 Expectant: 7.33 ± 0.06 No differences when stratified by parity 6) Shoulder dystocia: Induction: 5/108 Expectant: 6/109 p = not significant 7) Cephalohematoma: Induction: 6/134 (5 instrumental deliveries) Expectant: 3/139 (1 instrumental delivery) 8) Clavicular fracture: Induction: 0/134 Expectant: 2/139 9) Brachial plexus palsy: Induction: 0/134 Expectant: 2/139 10) Intraventricular hemorrhage: Induction: 44/134 had ultrasound; confirmed in 3 Expectant: 31/139 had ultrasound; confirmed in 2	

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Hannah, Hannah, Hellmann, et al., 1992 ***and*** **Goeree, Hannah, Hewson, 1995** **(cost-effectiveness analysis)**	Design: RCT, randomization stratified according to center, parity, and duration of gestation Interventions: 1) Induction of labor (n = 1701) Protocol: Subjects enrolled as outpatients. Labor to be induced within 4 days of randomization. If cervix < 3 cm dilated and < 50% effaced, and FHR normal, then patient given PGE_2 gel (0.5 mg) intracervically. Fetus monitored for minimum of 1 hour. Up to 3 doses of gel could be given at 6-hour intervals. If gel not used or did not induce labor, then labor induced by IV oxytocin, amniotomy, or both. Oxytocin infusion not started until 12 hours after last dose of gel. 2) Monitoring (n = 1706) Protocol: Subjects enrolled as outpatients and asked to do "kick counts" over 2-hour period each day, undergo NST 3 times per week, and undergo U/S assessments of AFV 2-3 times per week. If kick count < 6, then patients to contact physician and have NST within 12 hours. If NST nonreactive or showed deceleration in FHR, if AFV low (a pocket of < 3 cm), if obstetrical complications developed, or if gestational age reached 44 weeks, then fetus to be delivered either by	No. of subjects at start: 3418 Dropouts: 11 Loss to follow-up: NA No. of subjects at end: 3407 (Note: 7 of these 3407 women had infants with major congenital anomalies and were excluded from the analysis of perinatal and neonatal outcomes, as were 2 stillborns) Inclusion criteria: Live singleton fetus; ≥ 41 weeks gestation Exclusion criteria: Cervical dilatation ≥ 3 cm; gestational age ≥ 44 weeks; noncephalic presentation; lethal congential anomaly; diabetes mellitus; preeclampsia; intrauterine growth retardation; pre-labor rupture of membranes; need for urgent delivery; contraindications to vaginal delivery Age: Induction Monitoring < 20 4% 3% 20-35 86% 87% > 35 10% 10% Race: Induction Monitoring White 93% 92% Black 3% 3% Asian 2% 2% Other/ Unknown 2% 3% Gestational age at entry (in weeks): Induction Monitoring 40 3% 3%	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight > 4500 g 4) Shoulder dystocia 5) Meconium aspiration 6) Cord pH < 7.10 7) Admission to NICU 8) Stillbirths 9) Neonatal death 10) C-sections 11) Instrumental delivery 12) Length of stay 13) Hospital costs per patient 14) Professional fees per patient	1) Apgar score < 7 at 1 minute: Induction: 216/1700 (12.7%) Monitoring: 216/1698 (12.7%) p = not significant 2) Apgar score < 7 at 5 minutes: Induction: 18/1700 (1.1%) Monitoring: 20/1698 (1.2% p = not significant 3) Birthweight > 4500 g: Induction: 78/1700 (4.6%) Monitoring: 94/1698 (5.5%) p = not significant 4) Shoulder dystocia: Induction: 24/1701 (1.4%) Monitoring: 28/1706 (1.6%) p = not significant 5) Meconium aspiration: Induction: 96/1700 (5.7%) Monitoring: 95/1698 (5.6%) p = not significant 6) Cord pH < 7.10: Induction: 23/1700 (1.4%) Monitoring: 29/1698 (1.7%) p = not significant 7) Admission to NICU: Induction: 239/1700 (14.1%) Monitoring: 263/1698 (15.5%) p = not significant 8) Stillbirths: Induction: 0 Monitoring: 2 (no p-value reported) 9) Neonatal deaths: Induction: 0 Monitoring: 0	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Selection of mode of delivery was not standardized, but rather determined by the attending physician. For the cost analysis, minor costs were estimated from a sample of 129 charts. Sample size estimates based on reduction in incidence of Apgar score < 7 at 5 minutes. C-section rates higher among nulliparous women, older women, women with less dilatation at randomization, and women in "Black" and "Other" racial categories, independent of study group. Women induced in monitoring group less likely to receive prostaglandin for induction. *(continued on next page)*

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	inducing labor (using oxytocin or amniotomy) or by C-section. In every case, mode of delivery determined by attending physician. Dates: Nov 1985 - Dec 1990 Location: 22 sites "through-out Canada" (Canadian Multicentre Postterm Pregnancy Trial) Setting: 19 university hospitals and 3 community hospitals Type(s) of providers: Unspecified OB/GYN; radiologists Length of follow-up: None	41 88% 89% 42 9% 7% 43 < 1% < 1% Dating criteria: Either 1) LMP or known date of conception, confirmed by pregnancy test at < 6 weeks, physical exam at ≤ 20 weeks, or U/S at ≤ 26 weeks; *or* 2) U/S ≤ 26 weeks (if LMP uncertain); *or* 3) two consistent U/S at ≤ 26 weeks (if LMP unknown) Parity: 68% nulliparous (both groups) Bishop score: NR Other: Cervical dilatation before entry (in cm): Induction Monitoring 0 40% 40% 1-2 51% 49% 3-4 1% 1% Unknown 9% 10%		10) C-sections: *Overall:* Induction: 360 (21.2%) Monitoring: 418 (24.5%) p = 0.03 (controlled for parity, maternal age, cervical dilatation at time of randomization, and race) OR = 1.22 (95% CI, 1.02-1.45) *For fetal distress:* Induction: 97 (5.7%) Monitoring: 141 (8.3%) p = 0.003 11) Instrumental delivery: Induction: 473/1341 (35.3%) Monitoring: 449/1288 (34.9%) (no p-value reported) 12) Length of stay (mean): Induction: 3.9 days Monitoring: 4.0 days (no p-value reported) 13) Hospital costs (mean per patient in 1992 Canadian dollars): Induction: $2502 Monitoring: $2684 p < 0.0001 14) Professional fees (mean per patient in 1992 Canadian dollars): Induction: $437 Monitoring: $448 p = 0.025	

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Hedén, Ingemarsson, Ahlström, et al., 1991	Design: RCT, randomization by "birth registration number" Interventions: 1) Induction (n = 109) Protocol: Labor induced on day of recruitment by amniotomy and oxytocin infusion. (No further details provided.) 2) Monitoring ("expectant management") (n = 129) Protocol: Every-other-day clinical exam, cervical exam, and NST + weekly U/S assessment of AFV. If NST "ominous," then labor induced. If NST nonreactive, but not ominous, then oxytocin stress test (OST) performed. If OST normal, then monitoring protocol continued. If OST "ominous," then labor induced. If no pocket of fluid measuring at least 2 x 2 cm detected on U/S, then labor induced. Dates: NR; study conducted over a 3-year period Location: Lund and Ängelholm, Sweden Setting: University hospital and community hospital (2 sites) Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 238 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 238 Inclusion criteria: 42 weeks gestation; no complications; singleton fetus in vertex presentation; intact membranes; cervix < 4 cm; no regular contractions; normal NST; normal AFV Exclusion criteria: Prior C-section Age (mean ± SD): Induction, 29.5 ± 5.4; monitoring, 28.4 ± 4.9 Race: NR Gestational age at entry: 42 weeks (both groups) Dating criteria: U/S during weeks 16-18 Parity: Induction, 37% nulliparous; monitoring, 48% nulliparous Bishop score (mean ± SD): Induction, 5.3 ± 1.7; monitoring, 5.0 ± 2.1	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Severe dysmaturity 5) Admission to NICU 6) Meconium staining 7) C-sections 8) Forceps/vacuum extraction	1) Apgar score < 7 at 1 minute: Induction: 5/109 (4.6%) Monitoring: 6/129 (4.7%) p = not significant 2) Apgar score < 7 at 5 minutes: Induction: 3/109 (2.8%) Monitoring:1/129 (0.8%) p = not significant 3) Birthweight (mean): Induction: 4000 g Monitoring: 3900 g p = not significant 4) Severe dysmaturity: Induction: 4/109 (3.7%) Monitoring: 3/129 (2.3%) p = not significant 5) Admission to NICU: Induction: 10/109 (9.2%) Monitoring:8/129 (6.2%) p = not significant 6) Meconium staining: Induction: 15.6% Monitoring: 24.8% p = not significant 7) C-sections: Induction: 10/109 (9.2%) Monitoring: 9/129 (7.0%) p = not significant 8) Forceps/vacuum extraction: *Total:* Induction: 3/109 (2.8%) Monitoring: 20/129 (15.5%) p < 0.01 *For secondary arrest:* Induction: 2/109 (1.8%) Monitoring: 17/129 (13.2%) (p < 0.01)	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: - Interventions described: - Mode of delivery: - Sample size: - Statistical tests: - Gestational age: + Dating criteria: + Bishop score: + No sample size estimates. Unequal distribution of "semi-randomization" raises question of bias. Results not stratified by parity.

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Herabutya, Prasert-sawat, Tongyai, et al., 1992	Design: RCT, method of randomization not described Interventions: 1) Cervical ripening and induction (n = 57) Protocol: PGE_2 gel applied intracervically (6 tablets of 0.5 mg each mixed into 5 ml K-Y Jelly). Patient reassessed in 4-6 hours. If Bishop score > 6, then patient induced with amniotomy ± oxytocin (at discretion of obstetrician in charge of labor ward). If Bishop score < 6, then patient sent home, unless uterine contractions or "anticipated problem"; patients in latter categories kept in hospital and could receive 2nd dose after 6 hours if "urgent reasons" to repeat dose. Process repeated next morning, up to maximum of 3 doses. If Bishop score still < 6, then patient induced by amniotomy or oxytocin or both. 2) Monitoring (n = 51) Protocol: NST once weekly from 42-43 weeks and twice weekly after 43 weeks. Labor induced if NST abnormal, Bishop score > 6, or 44 weeks of gestation completed. For both groups, intrapartum management *not* dictated by study protocol. Dates: July 1987 - Jan 1991 Location: Bangkok, Thailand	No. of subjects at start: 108 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 108 Inclusion criteria: Gestational age ≥ 42 weeks; low risk Exclusion criteria: Bishop score > 6 Age (mean ± SD): Induction, 27.4 ± 4.1; monitoring, 27.1 ± 4.3 Race: 100% Thai Gestational age at entry: NR (required to be ≥ 42 weeks for entry into study) Dating criteria: LMP, with consistent obstetric exam at < 20 weeks Parity: Induction, 90% nulliparous; monitoring, 80% nulliparous Bishop score: NR (required to be ≤ 6 for entry into study)	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Meconium 4) Intubation required 5) Admission to NICU 6) Birthweight 7) Length of 1st stage of labor 8) C-sections 9) Instrumental deliveries	1) Apgar score < 7 at 1 minute: Induction: 15/57 (26.3%) Monitoring: 15/51 (29.4%) p = 0.89 2) Apgar score < 7 at 5 minutes: Induction: 1/57 (1.8%) Monitoring: 4/51 (7.8%) p = 0.19 3) Meconium: Induction: 8/57 (14.0%) Monitoring: 11/51 (21.6%) p = 0.44 4) Intubation required: Induction: 1/57 (1.8%) Monitoring: 4/51 (7.8%) p = 0.19 5) Admission to NICU: Induction: 1/57 (1.8%) Monitoring: 4/51 (7.8%) p = 0.19 6) Birthweight (mean ± SD): Induction: 3190 ± 429 g Monitoring: 3348 ± 421 g p = 0.06 7) Length of 1st stage of labor (mean ± SD): Induction: 8.15 ± 3.5 hours Monitoring: 9.15 ± 4.6 hours p =0.36 8) C-sections: *Overall:* Induction: 27/57 (47.4%) Monitoring: 24/51 (47.1%) p = 0.87 *For cephalopelvic disproportion:* Induction: 25/57 (43.9%)	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Results not stratified by parity. *(continued on next page)*

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Setting: University hospital Type(s) of providers: General OB/GYN Length of follow-up: None			Monitoring: 19/51 (37.3%) p = 0.62 *For fetal distress:* Induction: 2/57 (3.5%) Monitoring: 5/51 (9.8%) p = 0.26 9) Instrumental deliveries: Induction: 11/57 (19.3%) Monitoring: 9/51 (17.6%) p = 0.98	

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Katz, Yemini, Lancet, et al., 1983	Design: RCT, assignment to group by even/odd chart number Interventions: 1) Induction at 294 days (n = 78) Protocol: Labor induced by amniotomy and oxytocin infusion at 294 days. 2) Monitoring (n = 78) Protocol: Patients instructed to count fetal movements at home twice daily and to report to labor and delivery ward if movements decline by more than 50% or fall below 10 per hour. Patients seen every 3 days for assessment of "pelvic score" (Burnett, 1966), amnioscopy to check for meconium, OCT, and assessment of fetal movement count. If pelvic score > 4 or any of other 3 indicators "pathologic," then patient induced. Dates: NR Location: Jerusalem, Israel Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 156 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 156 Inclusion criteria: 294 days amenorrhea; "pelvic score" (Burnett, 1966) ≤ 4; vertex presentation; no obstetric pathology; no uterine scars; clear amniotic fluid by amnioscopy; normal NST; regular fetal movement perceived by mother Exclusion criteria: None specified Age (mean ± SD): Induction, 26.3 ± 4.1; monitoring, 26.5 ± 4.2 Race: NR Gestational age at entry: Both groups, 294 days Dating criteria: Positive pregnancy test within 6 weeks of LMP or 4 weeks following ovulation; *or* palpation of the uterus during 1st trimester and/or U/S before 30th week Parity: Induction, 46% primiparae; Monitoring, 45% primiparae Bishop score: NR; "pelvic score" (Burnett, 1966) required to be ≤ 4 for entry into study Other: NA	1) Apgar scores at 5 minutes (mean) 2) Apgar score < 7 at 5 minutes 3) Meconium staining 4) Intrapartum changes in FHR 5) Post-maturity syndrome 6) Birthweight (mean) 7) Birthweight > 4000 g 8) Perinatal death 9) C-sections 10) Duration of labor	1) Apgar scores at 5 minutes (mean): Induction: 9.5 Monitoring: 9.7 p = not significant 2) Apgar score < 7 at 5 minutes: Induction: 3/78 (3.8%) Monitoring: 1/78 (1.3%) (no p-value reported) 3) Meconium staining: Induction: 11/78 (14.1%) Monitoring: 12/78 (15.4%) p = not significant 4) Intrapartum changes in FHR: Induction: 9/78 (11.5%) Monitoring: 5/78 (6.4%) p = not significant 5) Post-maturity syndrome: Induction: 5/78 (6.4%) Monitoring: 11/78 (14.1%) p = not significant 6) Birthweight (mean): Induction: 3380 g Monitoring: 3540 g p = not significant 7) Birthweight > 4000 g: Induction: 6/78 (7.9%) Monitoring: 23/78 (29.5%) p < 0.05 8) Perinatal death: Induction: 1/78 (1.3%) Monitoring: 1/78 (1.3%) p = not significant 9) C-sections: Induction: 16/78 (20.5%) Monitoring: 7/78 (8.8%) p < 0.05	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: - Statistical tests: - Gestational age: + Dating criteria: + Bishop score: + Burnett, 1966 = Burnett JE. Preinduction scoring: an objective approach to induction of labour. Obstet Gynecol 1966;28:479-83. Results not stratified by parity. *(continued on next page)*

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				10) Duration of labor (mean ± SD): Induction: 9.4 ± 5.9 hours Monitoring: 6.7 ± 4.1 hours $p < 0.01$	

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Martin, Sessums, Howard, et al., 1989	Design: RCT, randomization by sealed envelope Interventions: 1) Induction (n = 12) Protocol: Patients admitted to hospital. Laminaria tent(s) inserted. Subsequently (usually the following morning), laminaria tents(s) removed, and labor induced by oxytocin infusion. Fetal heart tones monitored throughout labor. 2) Monitoring (n = 10) Protocol: Weekly monitoring, including U/S assessment of AFV, NST/CST, and cervical exam. Patients "admitted for delivery" if any monitoring test abnormal, or at the end of 43rd week of gestation. Dates: July 1987 - Jan 1988 Location: Jackson, MS Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 22 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 22 Inclusion criteria: Gestational age ≥ 41 weeks Exclusion criteria: Oligo-hydramnios (< 1 cm); nonreactive NST; positive CST; Bishop score > 5 Age (mean, with range): Induction, 23.3 (17-34); monitoring, 25.8 (18-37) Race: NR Gestational age at entry (mean, with range): Induction, 42 weeks (41-2/7 to 43-2/7); monitoring, 42 weeks (41-3/7 to 43-3/7) Dating criteria: LMP, 1st trimester pelvic exam, and/or U/S before 26 weeks Parity (mean): Induction, 0.76; monitoring, 0.58 Bishop score: NR	1) Apgar score at 1 minute 2) Apgar score at 5 minutes 3) Birthweight 4) Meconium 5) Complications 6) C-sections 7) Forceps-assisted deliveries 8) Length of labor 9) Maternal morbidity 10) Length of hospital stay	1) Apgar score at 1 minute (mean): Induction: 8.08 Monitoring: 8.4 p = not significant 2) Apgar score at 5 minutes (mean): Induction: 9.75 Monitoring: 9.7 p = not significant 3) Birthweight (mean, with range): Induction: 3560 g (2780-4110) Monitoring: 3472 g (2840-4180) p = not significant 4) Meconium: Induction: 1/12 (8%) Monitoring: 3/10 (30%) (no p-value reported) 5) Complications: Induction: 3/12 (25%) Monitoring: 1/10 (10%) (no p-value reported) 6) C-sections: Induction: 2/12 (17%) Monitoring: 1/10 (10%) p = not significant 7) Forceps-assisted deliveries: Induction: 3/12 (25%) Monitoring: 2/10 (25%) p = not significant 8) Length of labor (mean, with range): Induction: 6.33 hours (4-15) Monitoring: 8.3 hours (4-16) p = not significant 9) Maternal morbidity: Induction: 4/12 (33%) Monitoring: 2/10 (20%) (no p-value reported)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: - Results not stratified by parity. *(continued on next page)*

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				10) Length of hospital stay (mean, with range): Induction: 3.41 days (2-5) Monitoring: 2.6 days (2-6) p = not significant	

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units, 1994	Design: RCT, randomization by computer-generated random numbers Interventions: 1) PGE_2 gel + induction by oxytocin (n = 174) Protocol: PGE_2 gel (0.5 mg) inserted into intracervical canal within 24 hours of randomization. No repeat applications. FHR and uterine contractions monitored continuously for ≥ 4 hours. If no labor after 12 hours, then patient induced using amniotomy (where clinically feasible), followed by oxytocin infusion ("according to a uniform protocol"). If no active labor 24 hours after oxytocin infusion, then C-section performed or induction of labor continued. (Decision to perform C-section not dictated by study protocol.) 2) Placebo gel + induction by oxytocin (n = 91) Protocol: Same as in 1), above, except that placebo gel used instead of PGE_2 gel. 3) Monitoring (n = 175) Protocol: Weekly cervical exam + twice-weekly NST and U/S assessment of AFV. Spontaneous labor awaited, but labor could be induced if: Bishop score > 6; estimated fetal weight > 4500 g; medical or obstetric indication for delivery developed; largest pocket of amniotic fluid < 2	No. of subjects at start: 440 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 440 Inclusion criteria: Gestational age ≥ 287 days and < 301 days Exclusion criteria: Medical or obstetric complications requiring induction, C-section, or frequent monitoring; estimated fetal weight > 4500 g; Bishop score ≥ 7; non-reactive NST; amniotic fluid pocket < 2 cm Age (mean ± SD): PGE_2-oxytocin: 25.4 ± 5.7 Placebo-oxytocin: 25.4 ± 5.3 Monitoring: 26.1 ± 5.8 Race: PGE_2-oxytocin: 67% White, 32% Black, 1% not available Placebo-oxytocin: 63% White, 37% Black Monitoring: 60% White, 38% Black, 2% not available Gestational age at entry: PGE_2-oxytocin: 8l% 287-293 days; 19% 295-301 days Placebo-oxytocin: 79% 287-293 days; 21% 295-301 days Monitoring: 79% 287-293 days; 21% 295-301 days Dating criteria: Any one of following: 1) LMP + audible fetal heartbeat documented for ≥ 21 weeks by fetoscope or ≥ 30 weeks	1) Mechanical ventilation 2) Meconium aspiration 3) Nerve injury 4) Seizures 5) ≥ 1 adverse neonatal outcome 6) Apgar score < 4 at 5 minutes 7) Birthweight (mean) 8) Birthweight ≥ 4500 g 9) Time from randomization to delivery 10) Gestational age at delivery 11) Maternal infection 12) Maternal transfusion 13) Hyperstimulation 14) C-sections	1) Mechanical ventilation: PGE_2-oxytocin: 0 Placebo-oxytocin: 1/91 (1%) Monitoring: 1/175 (< 1%) (no p-value reported) 2) Meconium aspiration: PGE_2-oxytocin: 1/174 (< 1%) Placebo-oxytocin: 1/91 (1%) Monitoring: 2/175 (1%) (no p-value reported) 3) Nerve injury: PGE_2-oxytocin: 1/174 (< 1%) Placebo-oxytocin: 0 Monitoring: 0 (no p-value reported) 4) Seizures: PGE_2-oxytocin: 0 Placebo-oxytocin: 2/91 (2%) Monitoring: 1/175 (< 1%) (no p-value reported) 5) ≥ 1 adverse neonatal outcome: PGE_2-oxytocin: 1/174 (< 1%) Placebo-oxytocin: 3/91 (3%) Monitoring: 1/175 (< 1%) (no p-value reported) 6) Apgar score < 4 at 5 minutes: PGE_2-oxytocin: 0 Placebo-oxytocin: 0 Monitoring: 1/175 (< 1%) (no p-value reported) 7) Birthweight (mean ± SD): PGE_2-oxytocin: 3607 ± 382 g Placebo-oxytocin: 3532 ± 464 g Monitoring: 3606 ± 440 g (no p-value reported)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Sample size estimates based on perinatal morbidity/mortality and maternal mortality. *(continued on next page)*

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	cm; or abnormal NST followed by positive CST. If NST nonreactive, but CST negative, then testing repeated in 24 hours. Patients undelivered by 308 days (44 completed weeks) were released from the protocol and managed as "appropriate for the clinical situation." Dates: Dec 1987 - July 1989 Location: Multiple sites in US Setting: University hospitals Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	by Doppler; 2) LMP + compatible uterine size estimation at ≤ 24 weeks; 3) LMP + positive pregnancy test obtained early enough to assure that gestation exceeded 41 weeks; 4) if LMP uncertain, then fetal heartbeat documented for ≥ 32 weeks by Doppler; 5) U/S before 26 weeks Parity (% nulliparous): PGE_2-oxytocin: 60% Placebo-oxytocin: 59% Monitoring: 54% Bishop score (mean ± SD): PGE_2-oxytocin: 4.0 ± 1.4 Placebo-oxytocin: 3.8 ± 1.4 Monitoring: 3.9 ± 1.5		8) Birthweight ≥ 4500 g: PGE_2-oxytocin: 1/174 (< 1%) Placebo-oxytocin: 3/91 (3%) Monitoring: 6/175 (4%) (no p-value reported) 9) Time from randomization to delivery (median, with range): PGE_2-oxytocin: 36 hours (6-492) Placebo-oxytocin: 35 hours (7-487) Monitoring: 85 hours (5-538) p < 0.001 10) Gestational age at delivery: 287-293 days / 294-301 days / >302 days PGE_2-oxy: 64% / 34% / 1% Placebo-oxy: 66% / 32% / 2% Monitoring: 38% / 47% / 14% p < 0.001 11) Maternal infection: PGE_2-oxytocin: 33/174 (19%) Placebo-oxytocin: 13/91 (14%) Monitoring: 25/175 (14%) p = not significant 12) Maternal transfusion: PGE_2-oxytocin: 2/174 (1%) Placebo-oxytocin: 0 Monitoring: 3/175 (2%) p = not significant 13) Hyperstimulation: PGE_2-oxytocin: 2/174 (1%) Placebo-oxytocin: 1/91 (1%) Monitoring: 0 p = not significant 14) C-sections: PGE_2-oxytocin: 39/174 (22%) Placebo-oxytocin: 16/91 (18%) Monitoring: 32/175 (18%) p = not significant	

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Ohel, Rahav, Rothbart, et al., 1996	Design: RCT, assignment to group by even/odd registration number Interventions: 1) Induction (n = 70) Protocol: NST + U/S assessment of AFV performed before treatment. If NST normal, then 3-mg vaginal tablet of PGE_2 inserted into the posterior vaginal fornix. Patients sent home and instructed to return in 3-4 days for repeat testing and a further dose of PGE_2. 2) Monitoring (n = 104) Protocol: Patients "seen" twice weekly (monitoring protocol not described). Labor induced if patient passed 42 completed weeks of gestation. Dates: NR Location: Tiberias, Israel Setting: Unspecified hospital Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: 200 Dropouts: 26 Loss to follow-up: NA No. of subjects at end: 174 Inclusion criteria: Uncomplicated, singleton pregnancy; within 4 days after expected date of confinement Exclusion criteria: None specified Age (mean ± SD): Induction, 28.9 ± 4.0; monitoring, 28.2 ± 5.3 Race: NR Gestational age at entry: NR; *at delivery* (mean ± SD), Induction, 40.2 ± 0.5 weeks; monitoring, 40.9 ± 0.7 weeks Dating criteria: "Early" U/S Parity (mean ± SD): Induction, 2.2 ± 1.1; monitoring, 2.4 ± 1.5 Bishop score (mean ± SD): Induction, 4.1 ± 1.6; monitoring, 4.6 ± 1.6	1) Apgar scores at 5 minutes 2) Meconium staining 3) Birthweight > 4 kg 4) C-sections	1) Apgar scores at 5 minutes (mean ± SD): Induction: 9.5 ± 0.6 Monitoring: 9.4 ± 0.6 p = not significant 2) Meconium staining: Induction: 5/70 (7.1%) Monitoring: 20/104 (19.2%) p < 0.02 3) Birthweight > 4 kg: Induction: 6/70 (8.6%) Monitoring: 9/104 (8.7%) p = not significant 4) C-sections: Induction: 4/70 (5.7%) Monitoring: 6/104 (5.8%) p = not significant	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: + Interventions described: - Mode of delivery: - Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + 26 patients randomized to the induction group refused treatment and were excluded from analysis. Results not stratified by parity.

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Witter and Weitz, 1987	Design: RCT, randomization by computer-generated table of random numbers Interventions: 1) Induction at 42 weeks by oxytocin infusion + amniotomy (n = 103) Protocol: All patients instructed to keep 3-times-daily fetal motion charts. If decreased fetal motion, then OCT administered. If OCT positive, then patient delivered. If OCT negative, then patient continued with protocol. At 42 weeks, undelivered patients scheduled for induction of labor. Oxytocin infusion started at 7:00 AM with 1 mU/min and increased by 1 mU/min every 10 min until a dose of 30 mU/min reached or a regular pattern of adequate uterine contractions established. Amniotomy performed as soon as possible, but always after oxytocin had established regular contractions. If patient had intact membranes and was not in active phase labor by evening, the induction was stopped and the patient was rested overnight. The induction was restarted in the morning. If the patient failed to enter the active phase of labor by 20 hours of induction, then C-section performed. 2) Monitoring (principally by 24-hour urinary estriol	No. of subjects at start: 200 Dropouts: 5 (but included in analysis) Loss to follow-up: NA No. of subjects at end: 195 (200 included in analysis) Inclusion criteria: 41 completed weeks' gestation; uncomplicated pregnancy Exclusion criteria: None stated Age (mean ± SD): Induction, 20.95 ± 4.01; monitoring, 20.98 ± 3.67 Race: Induction, 20% White; monitoring, 34% White ($p < 0.05$) Gestational age at entry: NR; *at delivery* (mean ± SD), induction, 42.15 ± 1.92 weeks; monitoring, 42.41 ± 1.45 weeks Dating criteria: 2 or more of the following: certain LMP; basal body temperature indicating ovulation temperature shift for the present pregnancy; positive urinary pregnancy test at 6 weeks from LMP; fetal heart tones heard with DeLee stethoscope at 18-20 weeks; fundal height at the umbilicus at 20 weeks; fundal height in cm equal to gestational age in weeks within 2 cm from 20-34 weeks; early registration with dates equal to exam prior to 13 weeks; U/S dating by crown-rump length between 6 and 14 weeks or	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Small for gestational age 5) Large for gestational age 6) Post-maturity syndrome 7) Meconium aspiration 8) Endometritis 9) C-sections 10) Hospital stay	1) Apgar score < 7 at 1 minute: Induction: 20/103 (19.4%) Monitoring: 20/97 (21.1%) p = not significant 2) Apgar score < 7 at 5 minutes: Induction: 0 Monitoring: 2/97 (2.08%) p = not significant 3) Birthweight (mean ± SD): Induction: 3556.5 ± 436.3 g Monitoring: 3614.7 ± 472.2 g p = not significant 4) Small for gestational age: Induction: 0 Monitoring: 4/97 (4.43%) $p < 0.05$ 5) Large for gestational age: Induction: 21/103 (20.03%) Monitoring: 29/97 (29.59%) p = not significant 6) Post-maturity syndrome: Induction: 1/103 (0.97%) Monitoring: 2/97 (2.06%) p = not significant 7) Meconium aspiration: Induction: 2/103 (1.94%) Monitoring: 1/97 (1.03%) p = not significant 8) Endometritis: Induction: 12/103 (11.65%) Monitoring: 12/97 (12.37%) p = not significant 9) C-sections: *Overall:* Induction: 30/103 (29.13%) Monitoring: 27/97 (27.83%)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: - Results not stratified by parity. *(continued on next page)*

Evidence Table 2: Studies relevant to Key Question 2 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	creatinine ratio) (n = 97) Protocol: All patients instructed to keep 3-times-daily fetal motion charts. If decreased fetal motion, then OCT administered. If OCT positive, then patient delivered. If OCT negative, then patient continued with protocol. In addition, 24-hour urinary estriol creatinine ratio determined between 41 and 42 weeks. This increased to twice weekly at 42 completed weeks and three times weekly at 43 completed weeks. If 24-hour urinary estriol creatinine ratio ≤ 14 mg/g, then OCT performed. If OCT "reassuring," then patient kept as inpatient and given daily urinary estriol creatinine ratio tests and twice weekly OCTs until spontaneous labor occurred, or until delivery required (Bishop score ≥ 9 or signs of fetal compromise). If estriol creatinine ratio > 14 mg/g, the patient followed as outpatient until spontaneous labor occurred. Dates: NR Location: Baltimore, MD Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	by biparietal diameter prior to 26 weeks Parity: Induction, 51% nulliparous; monitoring, 41% nulliparous Bishop score: NR		p = not significant *For fetal distress:* Induction: 11/30 (36.67%) Monitoring: 13/27 (48.15%) p = not significant *For cephalopelvic disproportion/failure to progress:* Induction: 11/30 (36.67%) Monitoring: 13/27 (48.15%) p = not significant *For prolonged latent phase:* Induction: 7/30 (23/33%) Monitoring: 0 p < 0.01 *For breech presentation:* Induction: 1/30 (3.33%) Monitoring: 1/27 (3.70%) p = not significant 10) Hospital stay (mean ± SD): Induction: 4.74 ± 2.80 days Monitoring: 4.06 ± 1.90 days p < 0.05	

Evidence Table 3: Studies relevant to Key Question 3

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Allott and Palmer, 1993	Design: RCT, randomization by computer-generated list and sealed envelope Interventions: 1) Cervical exam to assess Bishop score + sweeping of the membranes (n = 99) Protocol: Examiner's index finger inserted as far as possible through internal cervical os and rotated twice through 360 degrees. Patients allowed to go home with a fetal movement chart. Instructed to telephone labor ward if they experienced decreased fetal movements, rupture of the membranes, or onset of labor. 2) Cervical exam to assess Bishop score alone (control) (n = 96) Protocol: Not described. Patients in both groups given deadline date for labor to be induced in the absence of spontaneous onset. Dates: NR (18-month period) Location: Reading, UK Setting: Community hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 195 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 195 Inclusion criteria: > 40 weeks gestation; no risk factors (e.g., IUGR or hypertension); able to introduce finger into cervix Exclusion criteria: None specified Age (mean ± SD): Sweeping, 27.7 ± 5.7; control, 27.5 ± 4.9 Race: NR Gestational age at entry (mean ± SD): Sweeping, 284.7 ± 3.3 days; control, 285.3 ± 3.5 days Dating criteria: Mid-trimester U/S Parity: Sweeping, 43% nulliparous; control, 46% nulliparous Bishop score: Both groups, 44% ≤ 6, 56% ≥ 7	1) Apgar score < 6 at 1 minute 2) Apgar score < 6 at 5 minutes 3) Serious neonatal infection 4) Antibiotics given 5) "Other serious neonatal outcome" 6) Induction of labor 7) C-sections 8) Epidural 9) Duration of labor 10) Precipitate labor (< 2 hours) 11) Time to delivery	1) Apgar score < 6 at 1 minute: Sweeping: 4/99 (4.0%) Control: 9/96 (9.4%) (no p-value reported) 2) Apgar score < 6 at 5 minutes: Sweeping: 0 Control: 0 (no p-value reported) 3) Serious neonatal infection: Sweeping: 0 Control: 1/96 (1%) (no p-value reported) 4) Antibiotics given: Sweeping: 0 Control: 1/96 (1%) (no p-value reported) 5) Other serious neonatal outcome: Sweeping: 0 Control: 0 (no p-value reported) 6) Induction of labor: Sweeping: 8/99 (8.1%) Control: 18/96 (18.8%) p = 0.035 7) C-sections: Sweeping: 4/99 (4.0%) Control: 5/96 (5.2%) p = not significant 8) Epidural: Sweeping: 19/99 (19.2%) Control: 20/96 (20.8%) p = not significant 9) Duration of labor (mean): Sweeping: 8.2 hours Control: 7.7 hours p = not significant	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Sample size estimates based on induction rates. Significant differences seen when results stratified by parity and Bishop score, except among primigravida with high Bishop score. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				10) Precipitate labor (< 2 hours): Sweeping: 14/99 (14.1%) Control: 19/96 (19.8%) p = not significant 11) Time to delivery (mean ± SEM): Sweeping: 2.24 ± 0.22 days Control: 5.18 ± 0.47 days p = 0.0001	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Atad, Hallak, Auslender, et al., 1996	Design: RCT, randomization by computer-generated list of random numbers Interventions: 1) PGE_2 (n = 30) Protocol: 3-mg tablet placed intravaginally. If contractions had not started or patient did not need analgesic agents 6 hours later, then second dose administered. If Bishop score still ≤ 4 at 12 hours, then patient treated with ARD. 2) Oxytocin (n = 30) Protocol: Oxytocin infusion given in initial dose of 1.5 mIU/min, with an increase of 1.5 mIU/min every 20 minutes until 3 contractions/10 minutes achieved. If Bishop score still ≤ 4 at 12 hours, then patient treated with ARD. 3) Atad Ripener Device (ARD) = double-balloon device invented by lead author (n = 35). Protocol: Device inserted into the cervix, and both balloons inflated with 100 ml or normal saline. Balloons deflated and device removed after 12 hours. If Bishop score still ≤ 4 at that time, then patient given PGE_2. Dates: NR Location: Haifa, Israel Setting: Community hospital	No. of subjects at start: 95 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 95 Inclusion criteria: Indication for induction; Bishop score ≤ 4; not in labor; singleton pregnancy; vertex presentation; intact membranes Exclusion criteria: Placenta previa; abnormal fetal monitoring; previous C-section Age (mean ± SD): PGE_2, 28.5 ± 5.2; oxytocin, 27.8 ± 5.7; ARD, 27.3 ± 4.2 Race: NR Gestational age at entry (mean ± SD): PGE_2, 38.8 ± 2.0 weeks; oxytocin, 39.6 ± 1.7 weeks; ARD, 40.0 ± 1.6 weeks Dating criteria: NR Parity: PGE_2, 57% primipara; oxytocin, 57 primipara; ARD, 54% primipara Bishop score (median, with range): 2 (0-4) all three groups Other: Indications for induction: Pregnancy-induced hypertension: 45% Postterm: 18% Diabetes mellitus: 7% Fetal growth restriction: 7%	1) Neonatal outcomes 2) Cervical dilation ≥ 3 cm at 12 hours 3) Failure of primary method 4) Time from induction to delivery 5) Success rate for vaginal delivery 6) C-sections	1) Neonatal outcomes: No quantitative data reported. Simply stated that neonatal outcome was "the same" for all 3 methods with respect to mean weight, Apgar scores at 1 and 5 minutes, and perinatal morbidity. 2) Cervical dilation ≥ 3 cm at 12 hours: PGE_2: 15/30 (50%) Oxytocin: 7/30 (23%) ARD: 30/35 (86%) $p < 0.01$ for ARD vs. PGE_2 and ARD vs. oxytocin 3) Failure of primary method: PGE_2: 6/30 (20%) Oxytocin: 16/30 (53%) ARD: 2/35 (6%) $p < 0.01$ for PGE_2 vs. oxytocin and ARD vs. oxytocin 4) Time from induction to delivery (mean ± SD): PGE_2: 23.2 ± 12.5 hours Oxytocin: 28.2 ± 14.7 hours ARD: 21.3 ± 7.0 hours p = not significant 5) Success rate for vaginal delivery: PGE_2: 21/30 (70%) Oxytocin: 8/30 (27%) ARD: 27/35 (77%) $p < 0.01$ for PGE_2 vs. oxytocin and ARD vs. oxytocin 6) C-sections: *Among patients successful with primary induction method:* PGE_2: 3/24 (13%) Oxytocin: 6/14 (43%) ARD: 6/33 (18%) $p < 0.05$ for PGE_2 vs. oxytocin and ARD vs. oxytocin	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (18% of total study population). Results not stratified by parity. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Type(s) of providers: Not specified Length of follow-up: None	Elective induction: 6% Nonreassuring NST: 6% Fetal death: 3% Other: 6%		*Among patients not successful with primary induction method:* PGE_2: 1/6 (17%) Oxytocin: 8/16 (50%) ARD: 1/2 (50%) (no p-value reported)	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Bell, Permezel, MacLennan, et al., 1993	Design: RCT, randomization by list of random numbers Interventions: 1) Relaxin gel (recombinant human, 1.5 mg) (n = 18) Protocol: Relaxin gel inserted into the posterior vaginal fornix on evening before scheduled induction. Patient remained recumbent for 1 hour. Spontaneous uterine activity, FHR, and maternal observations monitored overnight. If no labor after 15 hours, then induction protocol begun. This included surgical rupture of the membranes and IV administration of oxytocin at different dose schedules, according to the accepted regimen at each hospital. 2) Placebo gel (n = 22) Protocol: Same as above, except placebo gel used instead of relaxin. Dates: NR Location: Melbourne, Adelaide, and Clayton, Australia Setting: 4 hospitals of unspecified type Type(s) of providers: NR Length of follow-up: 6 weeks (relaxin levels and infant weight measured)	No. of subjects at start: 40 Dropouts: 0 Loss to follow-up: NR No. of subjects at end: NR (for 6-week follow-up) Inclusion criteria: Good maternal health, uncomplicated singleton pregnancy; gestational age 40-43 weeks; scheduled for induction for postdates pregnancy; cephalic presentation; unscarred uterus; maternal height > 1.5 m; normal blood pressure; no current medication Exclusion criteria: Abnormal placental location; antepartum hemorrhage; ruptured membranes; Calder score > 6 (modified Bishop score); fetal malformation; abnormal FHR tracing; IUGR; macrosomia; reduced AFV Age (mean ± SD): Relaxin, 25.7 ± 4.5; placebo, 27.3 ± 4.4 Race: NR Gestational age at entry (mean ± SD): Relaxin, 41.2 ± 0.4 weeks; placebo, 41.4 ± 0.7 weeks Dating criteria: NR Parity: Relaxin, 56% primiparas; placebo, 59% primiparas Bishop score: NR	1) Stillbirths 2) Neonatal deaths 3) Abnormal FHR tracings warranting intervention 4) Apgar scores at 1, 5, and 10 minutes 5) Cord blood gases 6) Birthweight 7) Forceps-assisted deliveries 8) C-sections 9) Time to delivery 10) Duration of labor	1) Stillbirths: None in either group. 2) Neonatal deaths: None in either group. 3) Abnormal FHR tracings warranting intervention: Relaxin: 7/18 (39%) Placebo: 7/22 (32%) p = not significant 4) Apgar scores at 1, 5, and 10 minutes: No statistically significant differences between two groups (no quantitative data reported) 5) Cord blood gases: No statistically significant differences between two groups (no quantitative data reported) 6) Birthweight (mean ± SD): Relaxin: 3634 ± 403 g Placebo: 3673 ± 310 g p = 0.73 7) Forceps-assisted deliveries Relaxin: 6/18 (33.3%) Placebo: 6/22 (27.3%) p = not significant 8) C-sections: Relaxin: 2/18 (11.1%) Placebo: 4/22 (18.2%) p = not significant 9) Time to delivery (mean ± SD): Relaxin: 23.6 ± 4.8 hours Placebo: 24.8 ± 4.8 hours p = 0.33 10) Duration of labor (mean ± SD): Relaxin: 7.1 ± 3.4 hours Placebo: 7.5 ± 3.4 hours *(continued on next page)*	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + First trial ever conducted of recombinant human relaxin in pregnant women. Low dose used deliberately. Primarily interested in establishing safety in pregnant women. Results not stratified by parity.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
		Other: Calder score: Score Relaxin Placebo ≤ 4 33% 32% 5 50% 41% 6 17% 27%		p = 0.49	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Berghella, Rogers, and Lescale, 1996	Design: RCT, randomization by computer-generated random number table and sealed envelopes Interventions: 1) Stripping of the membranes (n = 73) Protocol: Stripping of the membranes performed weekly starting at 38 weeks by separating an approximately 2-3-cm section the lower membranes from its cervical attachment with at least two circumferential passes of the index finger. 2) Cervical exam (control) (n = 69) Protocol: "Gentle cervical examination" performed weekly starting at 38 weeks. Dates: Jul - Oct 1991 and Jul - Oct 1993 Location: New York, NY Setting: Outpatient clinic/ physician office Type(s) of providers: General OB/GYN Length of follow-up: NA	No. of subjects at start: 149 Dropouts: 7 (excluded at 38 weeks due to long, closed cervices not amenable to stripping) Loss to follow-up: NA No. of subjects at end: 142 Inclusion criteria: First presented to clinic at gestational age ≤ 20 weeks Exclusion criteria: Multiple pregnancy; placenta previa; low-lying placenta; nonvertex presentation; IUGR; any medical complication of pregnancy; long, closed cervix not amenable to stripping at time of intervention (38 weeks) Age (mean ± SD): Stripping, 27.19 ± 6.1; control, 27.12 ± 5.6 Race: 100% Asian Gestational age at entry: 38 weeks Dating criteria: Pelvic exam during first 12 menstrual weeks to confirm size appropriate for dates and/or U/S before 20th week Parity: Stripping, 48% nulliparas; control, 62% nulliparas (p = not significant) Bishop score (mean ± SD): Stripping, 3.49 ± 2.7; control, 2.46 ± 2.3	1) Delivery after 41 weeks 2) Vacuum-assisted delivery 3) Forceps-assisted delivery 4) C-sections 5) Days to delivery (overall and broken down by Bishop score and parity)	1) Delivery after 41 weeks: Stripping: 4/73 (5%) Control: 15/69 (22%) p < 0.01 2) Vacuum-assisted delivery: Stripping: 2/73 (3%) Control: 3/69 (4%) p = not significant 3) Forceps-assisted delivery: Stripping: 5/73 (7%) Control: 4/69 (6%) p = not significant 4) C-sections: Stripping: 0/73 Control: 3/69 (4%) p = not significant 5) Days to delivery (mean ± SD): *Overall:* Stripping: 8.2 ± 6.3 Control: 12.2 ± 7.1 p < 0.002 *Broken down by Bishop score:* Bishop score ≤ 3: Stripping (n =39): 8.6 ± 6.4 Control (n = 44): 12.5 ± 6.8 p ≤ 0.02 Bishop score > 3: Stripping (n = 34): 6.5 ± 5.4 Control (n = 25): 11.5 ± 8.2 p = 0.10 *Broken down by parity:* Nulliparas: Stripping (n = 35): 7.8 ± 6.0 Control (n = 43): 12.9 ± 6.6 p < 0.09 Multiparas: Stripping (n = 38): 7.2 ± 5.9	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Sample size estimates based on proportion of patients delivering at ≥ 41 weeks. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				Control (n = 26): 11.0 ± 7.9 p = 0.10	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Boulvain, Fraser, Marcoux, et al., 1998	Design: RCT, randomization by computer-generated list of random numbers and sealed envelopes Interventions: 1) Sweeping of the membranes (n = 99) Protocol: Sweeping performed using circular movements of examining finger between the lower segment of the uterus and the fetal membranes. If membranes could not be reached, then examiner attempted to dilate cervix manually. If successful, then sweeping performed; if not, then cervical massage performed. 2) Control (n = 99) Protocol: Vaginal exam performed for Bishop scoring only In both groups, post-intervention management, including method of induction and intrapartum interventions, were left to the discretion of the treating obstetrician. Dates: Apr 1995 - Oct 1996 Location: 3 sites in the province of Quebec, Canada Setting: 3 university hospitals Type(s) of providers: General OB/GYN	No. of subjects at start: 200 Dropouts: 2 Loss to follow-up: NA No. of subjects at end: 198 Inclusion criteria: Medical indication for nonurgent induction; gestational age ≥ 266 days; single fetus; cephalic presentation Exclusion criteria: None specified Age (mean ± SD): Sweeping, 28.5 ± 5.5; control, 29.2 ± 4.6 Race: NR Gestational age at entry (mean ± SD): Sweeping, 281.9 ± 5.0 days; control, 281.5 ± 4.5 days Dating criteria: LMP plus 2nd trimester U/S Parity: Sweeping, 58% nulliparous; control, 49% nulliparous Bishop score (mean ± SD): Sweeping, 5.8 ± 2.2; control, 5.3 ± 2.3 Other: Indications for induction: Postterm (> 287 days): 85% Hypertension: 4% Diabetes: 2.5% IUGR: 1.5% Other: 7%	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Admission to NICU 5) Neonatal infection 6) Cephalhematoma 7) Convulsions 8) Respiratory distress 9) Induction of labor 10) Fever during labor or postpartum 11) Forceps/vacuum delivery 12) C-sections 13) Time from randomization to onset of labor	1) Apgar score < 7 at 1 minute: Sweeping: 5/99 Control: 8/99 p = 0.40 2) Apgar score < 7 at 5 minutes: Sweeping: 3/99 Control: 0/99 p = 0.25 3) Birthweight (mean ± SD): Sweeping: 3501 ± 436 g Control: 3633 ± 438 g p = 0.04 4) Admission to NICU: Sweeping: 6/99 Control: 6/99 p = 1.00 5) Neonatal infection: Sweeping: 1/99 Control: 1/99 p = 1.00 6) Cephalhematoma: Sweeping: 5/99 Control: 2/99 p = 0.44 7) Convulsions: Sweeping: 1/99 Control: 0/99 p = 1.00 8) Respiratory distress: Sweeping: 0/99 Control: 1/99 p = 1.00 9) Induction of labor: Sweeping: 49/99 Control: 59/99 p = not significant	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (85% of total study population). Positive effect in multiparas with Bishop score > 6 (RR, 0.55; 95% CI, 0.31-0.98), but not in other groups. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Length of follow-up: None			10) Fever during labor or postpartum: Sweeping: 8/99 Control: 8/99 p = not significant 11) Forceps/vacuum delivery: Sweeping: 36/99 Control: 27/99 (no p-value reported) 12) C-sections: Sweeping: 12/99 Control: 12/99 p = 0.37 13) Time from randomization to onset of labor (mean): Sweeping: 76 hours Control: 98 hours p = 0.01	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Brennand, Calder, Leitch, et al., 1997	Design: RCT, randomized by computer-generated list Interventions: 1) 4 mg recombinant human relaxin (n = 25) given between 37 and 42 weeks gestation. Gel introduced into posterior fornix; NST monitored for 4 hours post-treatment, then every 4 hours for 24 hours or until delivery. 2) 2 mg relaxin (n = 25), given in same manner 3) 1 mg relaxin (n = 23), given in same manner 4) Placebo gel (n = 23), given in same manner In all groups, induction started by placing 2 mg PGE_2 gel intravaginally 15 hours after relaxin, amniotomy ± additional PGE_2 Dates: NR Location: Edinburgh, Glasgow, Manchester, and Oxford, UK Setting: University hospitals Providers: Unspecified OB/GYN Length of follow-up: None	No of subjects at start: 96 Drop-outs: 0 Loss to follow-up: NA No of subjects at end: 96 Inclusion criteria: Gestational age ≥ 37 weeks, Bishop score ≤ 4 Exclusion criteria: Uterine scar; ruptured membranes; evidence of placental abruption or previa; systemic disease; recent ingestion of NSAIDs; fetal malformation; abnormalities in fetal growth, size, or amniotic fluid volume Age (mean): 4 mg: 25.8 2mg: 26.7 1 mg: 26.8 Placebo: 27.0 Race: NR Gestational age at entry: 4 mg: 40.1 weeks 2 mg: 39.9 1 mg: 39.6 Placebo: 40.0 Dating criteria: NR Parity (% nulliparous): 4 mg: 76% 2 mg: 88% 1 mg: 87% Placebo: 78% Bishop score (mean): 4 mg: 2.5 2 mg: 2.8	1) Change in Bishop score between baseline and 15 hours 2) Spontaneous labor 3) Treatment to delivery 4) Cesarean delivery 5) Perinatal morbidity/mortality	1) Change in Bishop score between baseline and 15 hours: 4 mg: 1.32 2 mg: 1.76 1 mg: 1.36 Placebo: 1.64 p = 0.85 2) Spontaneous labor: 4 mg: 2/25 2 mg: 5/25 1 mg: 1/23 Placebo: 2/23 p = 0.93 3) Treatment to delivery (mean): 4 mg: 36.7 hours 3 mg: 39.3 hours 1 mg: 29.9 hours Placebo: 28.0 hours p = 0.31 3) Cesarean delivery: 4 mg: 4/25 3 mg: 8/25 1 mg: 3/23 Placebo: 4/23 p = 0.45 4) Perinatal morbidity/mortality: No deaths in any group. No significant differences reported except higher baseline fetal heart rates in all relaxin groups compared to placebo.	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy. Study underpowered to detect differences in important outcomes. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
		1 mg: 3.0 Placebo: 2.9 Other: Indications for induction: "Most" pregnancy-induced hypertension or prolonged pregnancy; numbers not given			

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Buser, Mora, and Arias, 1997	Design: RCT, randomization by random numbers table and sealed envelopes Interventions: 1) Misoprostol (n = 76) Protocol: 50-µg tablet placed in posterior vaginal fornix using a speculum. Dose repeated every 4 hours until patient developed an adequate contraction pattern (≥ 3 contractions in 10 minutes), cervix reached ≥ 3 cm dilation and 100% effacement, or SROM occurred. Maximum of 3 doses. Oxytocin augmentation started 4 hours after last dose if adequate pattern of contraction still not obtained. 2) PGE_2 (n = 79) Protocol: PGE_2 gel (0.5 mg) administered intracervically using a speculum. Dose repeated every 6 hours until patient developed an adequate contraction pattern (≥ 3 contractions in 10 minutes), cervix reached ≥ 3 cm dilation and 100% effacement, or SROM occurred. Maximum of 3 doses. Oxytocin augmentation started 6 hours after last dose if adequate pattern of contraction still not obtained. Dates: July 1994 - Dec 1995 Location: St. Louis, MO	No. of subjects at start: 155 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 155 Inclusion criteria: Admitted for induction; singleton pregnancy at term; cephalic presentation; reassuring FHR tracing; Bishop score ≤ 5 Exclusion criteria: Ruptured membranes; low-lying placenta; partial or complete placenta previa; prior C-section; parity ≥ 6; strong clinical suspicion of fetopelvic disproportion; history of asthma, glaucoma, or cardiac disease Age (mean ± SD): Misoprostol, 27.7 ± 5.6; PGE_2, 27.1 ± 5.8 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 39.2 ± 1.9 weeks; PGE_2, 39.3 ± 1.8 weeks Dating criteria: NR Parity: Misoprostol, 84% nulliparas; PGE_2, 82% nulliparas Bishop score (mean ± SD): Misoprostol, 2.66 ± 1.3; PGE_2, 2.64 ± 1.4 Other: Indications for induction: Postterm: 35% Preeclampsia: 28%	1) Apgar score < 6 at 5 minutes 2) Birthweight 3) Admission to NICU 4) Number of days in NICU 5) Nonreassuring FHR tracing with hyper-stimulation 6) Change in Bishop score 7) Time from induction to delivery 8) C-sections 9) Spontaneous vaginal delivery	1) Apgar score < 6 at 5 minutes: Misoprostol: 2/76 (3%) PGE_2: 0/79 p = not significant 2) Birthweight (mean ± SD): Misoprostol: 3435 ± 564 g PGE_2: 3383 ± 618 g p = not significant 3) Admission to NICU: Misoprostol: 7/76 (9%) PGE_2: 0/79 p = not significant 4) Number of days in NICU (mean): Misoprostol: 14 days PGE_2: 13 days p = not significant 5) Nonreassuring FHR tracing with hyper-stimulation: Misoprostol: 14/76 (18%) PGE_2: 0/79 p < 0.001 6) Change in Bishop score (mean ± SD): Misoprostol: 3.53 ± 2.1 PGE_2: 2.7 ± 1.8 p = 0.01 7) Time from induction to delivery (mean ± SD): Misoprostol: 15.8 ± 7.0 hours PGE_2: 24.2 ± 11.0 hours p < 0.01 8) C-sections: *Overall:* Misoprostol: 27/76 (36%) PGE_2: 17/79 (22%) p = not significant	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (35% of total study population, unevenly distributed: 41% of miso-prostol group, 29% of PGE_2 group [p = not significant, but study underpowered]). Sample size estimates based on change in Bishop score, active labor, and C-section rate. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Setting: Community hospital Type(s) of providers: Not specified Length of follow-up: None	Decreased amniotic fluid: 10% Large for gestational age: 10% Gestational diabetes: 3% Fetal growth restriction: 3% Other: 11%		*For nonreassuring FHR tracing:* Misoprostol: 19/76 (25%) PGE_2: 4/79 (5%) $p < 0.001$ 9) Spontaneous vaginal delivery: Misoprostol: 25/76 (33%) PGE_2: 37/79 (47%) p = not significant	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Buttino and Garite, 1990	Design: RCT, randomization performed by dispensing pharmacy Interventions: 1) PGE_2 gel (0.5 mg) (n = 23) Protocol: Patient underwent CST/NST, which had to be negative/reactive before treatment administered. PGE_2 gel placed intra-cervically using a syringe. Patient observed on external fetal monitor for 1 hour and then allowed to go home. 2) Placebo (n = 20) Protocol: Same as above, except that placebo gel used in place of PGE_2. Dates: NR Location: Long Beach, CA Setting: Unspecified hospital Type(s) of providers: NR Length of follow-up: NA	No. of subjects at start: 43 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 43 Inclusion criteria: Gestational age ≥ 41-6/7 weeks (279 days); no contraindications to prostaglandins Exclusion criteria: None stated Age (mean): PGE_2, 24.9; placebo, 25.8 Race: NR Gestational age at entry (mean): PGE_2, 42.3 weeks; placebo, 42.5 weeks Dating criteria: Any two of the following: LMP; 1st trimester pelvic exam consistent with dates; U/S demonstrating either a crown-rump length at 6-11 weeks or biparietal diameter and femur measurements at 17-20 weeks consistent with dates Parity: PGE_2, 43% primigravidas; placebo, 30% primigravidas (p = not significant) Bishop score (mean ± SD): PGE_2, 2.8 ± 0.8; placebo, 2.2 ± 1.3	1) Apgar scores at 1 minute 2) Apgar scores at 5 minutes 3) Birthweight 4) Time to delivery 5) Duration of labor 6) Change in Bishop score 7) C-sections	1) Apgar scores at 1 minute (mean ± SD): PGE_2: 7.8 ± 1.1 Placebo: 8.2 ± 0.8 p = not significant 2) Apgar scores at 5 minutes (mean ± SD): PGE_2: 8.9 ± 0.3 Placebo: 9.0 ± 0.2 p = not significant 3) Birthweight (mean ± SD): PGE_2: 3644.6 ± 416.7 g Placebo: 3840.8 ± 574.4 p = not significant 4) Time to delivery (mean ± SD): PGE_2: 311.2 ± 244.8 hours Placebo: 379.6 ± 186.7 hours p = not significant 5) Duration of labor (mean ± SD): PGE_2: 10.6 ± 6.9 hours Placebo: 9.0 ± 4.2 hours p = not significant 6) Change in Bishop score (mean ± SD): PGE_2: 3.8 ± 2.3 Placebo: 3.0 ± 2.3 p = not significant 7) C-sections: PGE_2: 5/23 (21.7%) Placebo: 7/20 (35.0%) p = not significant	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: - Statistical tests: - Gestational age: + Dating criteria: + Bishop score: + Underpowered to detect differences either at baseline or at outcome time points. Results not stratified by parity.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Cammu and Haitsma, 1998	Design: RCT, randomization by computer-generated list of random numbers and sealed envelopes Interventions: 1) Sweeping of the membranes (n = 140) Protocol: Sweeping of the membranes performed weekly beginning at 39 completed weeks. Digital separation of 2-3 cm of the membranes from the lower uterine segment performed, rotating the finger at least twice through 360 degrees. Closed cervix stretched digitally until membrane sweeping could be carried out. Closed cervix that would not admit a finger was vigorously massaged. 2) Control (n = 138) Protocol: Routine pelvic exam performed weekly beginning at 39 completed weeks. In both groups, induction planned from 41 completed weeks onward and performed according to standard protocol (amniotomy ± oxytocin, with cervical ripening beforehand, if necessary). Dates: NR (patients enrolled over a 25-month period) Location: Brussels, Belgium Setting: Antenatal clinic of university hospital	No. of subjects at start: 287 Dropouts: 9 Loss to follow-up: NA No. of subjects at end: 278 Inclusion criteria: Gestational age 39 weeks; nulliparous; singleton fetus; cephalic presentation; no risk factors Exclusion criteria: None specified Age (mean ± SD): Sweeping, 27.6 ± 3.8; control, 27.6 ± 4.0 Race: NR; clinic said to serve "mostly urban middle class Caucasian women" Gestational age at entry (mean ± SD): 273.3 ± 2.4 days; 273.2 ± 2.5 days Dating criteria: U/S (not specified whether 1st or 2nd trimester) Parity: 100% nulliparous Bishop score (mean ± SD): Sweeping, 3.35 ± 1.8; control, 3.39 ± 1.6	1) Apgar score < 7 at 5 minutes 2) Arterial cord blood pH < 7 3) Birthweight 4) Gestational age at delivery 5) Induction of labor 6) Instrumental delivery 7) C-sections 8) Time from randomization to delivery	1) Apgar score < 7 at 5 minutes: Sweeping: 3/140 (2%) Control: 5/138 (4%) p = 0.490 2) Arterial cord blood pH < 7: Sweeping: 7/140 (5%) Control: 8/138 (6%) p = 0.976 3) Birthweight (mean ± SD): Sweeping: 3400 ± 375 g Control: 3459 ± 411 g p = not significant 4) Gestational age at delivery: *Mean ± SD:* Sweeping: 282.8 ± 5 days Control: 283.8 ± 6 days p = not significant *Percentage > 287 days:* Sweeping: 27/140 (19%) Control: 45/138 (33%) OR = 0.49 (95% CI, 0.29-0.86) 5) Induction of labor: Sweeping: 15/140 (11%) Control: 36/138 (26%) OR = 0.34 (95% CI, 0.18-0.66) 6) Instrumental delivery: Sweeping: 23/140 (16%) Control: 18/138 (13%) OR = 1.31 (95% CI, 0.67-2.55) 7) C-sections: Sweeping: 5/140 (4%) Control: 8/138 (6%) OR = 0.60 (95% CI, 0.19-1.89) 8) Time from randomization to delivery (mean ± SD): Sweeping: 9.4 ± 5 days	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + 24/140 women in the membrane-sweeping group (17%) had cervixes inaccessible to an examining finger and received cervical massage only. These women were not excluded from the analysis. Sample size estimates based on proportion of patients reaching 41 weeks. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Type(s) of providers: NR Length of follow-up: None			Control: 10.6 ± 6 days (no p-value reported)	
Chang and Chang, 1997	Design: RCT, method of randomization not described Interventions: 1) PGE_2 (n = 30) Protocol: 3-mg tablet placed in posterior vaginal fornix. Dose repeated every 6 hours until satisfactory uterine activity achieved. Maximum dose permitted was 9 mg. 2) Misoprostol (n = 30) Protocol: 50-µg tablet placed in posterior vaginal fornix. Dose repeated every 4 hours until satisfactory uterine activity achieved. Maximum dose permitted was 600 µg. In both groups, oxytocin augmentation initiated if Bishop score ≥ 9, but uterine contractions inadequate (< 3 per 10 minutes). Dates: July 1994 - June 1995 Location: Tainan, Taiwan Setting: University hospital Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: Dropouts: Loss to follow-up: No. of subjects at end: Inclusion criteria: Scheduled for induction; term singleton pregnancy; Bishop score ≤ 5; no regular uterine contractions Exclusion criteria: Contra-indications to vaginal prostaglandins; any maternal illness for which induction of labor not appropriate Age (mean ± SD): PGE_2, 28.9 ± 5.3; misoprostol, 27.6 ± 6.7 Race: NR Gestational age at entry (mean ± SD): PGE_2, 39.3 ± 2.4 weeks; misoprostol, 38.9 ± 3.1 weeks Dating criteria: NR Parity: 100% nulliparous in both groups Bishop score (mean ± SD): PGE_2, 4.3 ± 1.1; misoprostol, 4.2 ± 0.5 Other: Indications for induction: Excess maternal weight gain (> 16 kg): 42% Postterm: 40% Hypertension: 18%	1) Apgar scores < 7 at 1 and 5 minutes 2) Birthweight 3) Cord arterial pH 4) Time from induction to delivery 5) Hyperstimulation 6) Vacuum extractions 7) C-sections	1) Apgar scores < 7 at 1 and 5 minutes: No quantitative data reported. Simply stated that proportion of neonates with Apgar ≤ 7 at 1 and 5 minutes was "the same" in both groups. 2) Birthweight (mean ± SD): PGE_2: 3376 ± 432 g Misoprostol: 3285 ± 580 g p = not significant 3) Cord arterial pH (mean ± SD): PGE_2: 7.32 ± 0.91 Misoprostol: 7.29 ± 0.73 p = not significant 4) Time from induction to delivery (mean ± SD): PGE_2: 25.7 ± 3.8 hours Misoprostol: 16.5 ± 2.7 hours p < 0.001 5) Hyperstimulation: PGE_2: 8.9% Misoprostol: 13.4% p < 0.05 6) Vacuum extractions: PGE_2: 6% Misoprostol: 10% p = not significant 7) C-sections: PGE_2: 6% Misoprostol: 10% p = not significant	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: - Interventions described: + Mode of delivery: - Sample size: - Statistical tests: - Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (40% of total study population).

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Chatterjee, Ramchandran, Ferlita, et al., 1991	Design: RCT, randomization by card shuffling Interventions: 1) 2 mg PGE_2 gel applied in posterior fornix (n = 15) 12 hours prior to induction with oxytocin 2) Placebo gel (n = 18) In both groups, second application possible if induction unsuccessful. Dates: Jul 1983 - Apr 1984 Location: Newark, NJ Setting: University hospital Providers: Unspecified OB/GYN Length of follow-up: None	No of subjects at start: 38 Dropouts: 0 Loss to follow-up: NA No of subjects at end: 38 Inclusion criteria: NR Exclusion criteria: NR Age (mean ± SD): PGE_2: 24.2 ± 1.1 Placebo: 25.1 ± 1.3 Race: NR Gestational age at entry: PGE_2: 39.1 ± 0.5 Placebo: 38.4 ± 0.9 Dating criteria: NR Parity: NR Bishop score: NR Other: 18% induced for prolonged pregnancy	1) Change in Bishop score 2) Cesarean section 3) Mean Apgar score at 1 minute 4) Mean Apgar score at 5 minutes	1) Change in Bishop score: Data presented graphically; statistically significant greater change with PGE_2 ($p < 0.01$). 2) Cesarean section: PGE_2: 7/15 Placebo: 5/18 3) Mean Apgar score at 1 minute: PGE_2: 6.8 Placebo: 6.8 4) Mean Apgar score at 5 minutes: PGE_2: 7.9 Placebo: 8.1	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: +

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Chayen, Tejani, and Verma, 1986	Design: RCT, allocation to treatment group by even/odd hospital ID number Interventions: 1) Nipple stimulation using breast pump (n = 30) Protocol: Patients admitted to labor ward, placed on an external monitor, and assigned a Bishop score. Vaseline applied to nipple. Breast pump turned on to normal setting (250 mmHg of negative pressure). Pump alternated from right to left breast every 15 minutes. Once regular contractions occurred and cervix ≥ 2 cm dilated, then patient underwent amniotomy and had internal pressure catheter placed. If active phase not reached or active phase arrested, then patient switched to oxytocin protocol. 2) Induction using oxytocin (control) (n = 32) Protocol: Patients admitted to labor ward, placed on an external monitor, and assigned a Bishop score. Induction initiated with 2 µm/min of oxytocin, with gradual increments until "adequate uterine activity" (≥ 200 Montevideo units) achieved. Once regular contractions occurred and cervix ≥ 2 cm dilated, then patient underwent amniotomy and had internal pressure catheter placed. Patients who	No. of subjects at start: 62 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 62 Inclusion criteria: Admitted for induction of labor Exclusion criteria: None specified Age: NR Race: NR Gestational age at entry (mean ± SD): Breast pump, 39.31 ± 2.33 weeks, 9/30 (30%) "postdates"; oxytocin, 40.18 ± 1.90 weeks, 8/32 (25%) "postdates" Dating criteria: NR Parity: Breast pump, 43% nulliparous; oxytocin, 53% nulliparous Bishop score (mean ± SD): Breast pump, 5.48 ± 1.87; oxytocin, 6.62 ± 1.77 (p = 0.05) Other: Indications for induction: Preeclampsia: 44% Postterm: 29% Other: 27%	1) Failure to reach active phase 2) Time to regular contractions 3) Time to adequate labor 4) Time to active phase 5) C-sections	1) Failure to reach active phase: Breast pump: 3/30 (10%) Oxytocin: 4/32 (12.5%) p = not significant 2) Time to regular contractions (mean ± SD): Breast pump: 5.68 ± 6.13 minutes Oxytocin: 61.55 + 42.62 minutes p = 0.0005 3) Time to adequate labor (mean ± SD): Breast pump: 1.52 ± 1.075 hours Oxytocin: 3.41 ± 2.22 hours p = 0.0005 4) Time to active phase (mean ± SD): Breast pump: 4.84 ± 3.33 hours Oxytocin: 6.90 ± 4.21 hours p = 0.05 5) C-sections: Breast pump: 8/30 (26.7%) Oxytocin: 14/32 (43.7)% p = not significant	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: - Interventions described: + Mode of delivery: - Sample size: - Statistical tests: - Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (29% of total study population). Significant difference in baseline Bishop scores – bias in favor of oxytocin. Results not stratified by parity. Study underpowered to detect difference at baseline or in outcomes. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	failed induction delivered by C-section. Dates: NR Location: Stony Brook, NY Setting: University hospital; community hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None				

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Chuck and Huffaker, 1995	Design: RCT, randomization by computer and sealed envelopes Interventions: 1) Misoprostol (n = 49) Protocol: 50-µg tablet placed in posterior vaginal fornix. Additional doses given every 4 hours for a maximum of 5 doses. 2) PGE_2 (n = 50) Protocol: Gel (0.5 mg) placed intracervically. Additional doses given every 4 hours for a maximum of 5 doses. In both groups, dosing halted for hyperstimulation or if patient having ≥ 3 contractions/10 minutes. Oxytocin used if no labor after maximum dose or if labor progress arrested for > 2 hours. AROM performed when cervix > 3 cm. Dates: Sep 1993 - Jan 1994 Location: Los Angeles, CA Setting: Community hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 103 Dropouts: 4 (excluded from analysis due to protocol violations) Loss to follow-up: NA No. of subjects at end: 99 Inclusion criteria: Gestational age 35-42 weeks; admitted for induction of labor Exclusion criteria: Nonvertex presentation; uterine scar other than from prior low-transverse C-section; ominous FHR tracing; multiple gestation; complete cervical effacement Age (mean ± SD): Misoprostol, 29.3 ± 6.7; PGE_2, 28.7 ± 6.4 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 29.7 ± 1.7 weeks; PGE_2, 39.7 ± 1.3 weeks Dating criteria: NR Parity (mean ± SD): Misoprostol, 0.8 ± 0.9 (52% nulliparous); PGE_2, 0.8 ± 0.9 (48% nulliparous) Bishop score: Misoprostol, 53% ≤ 3; PGE_2, 52% ≤ 3 Other: Indications for induction: PROM: 28% Postterm: 18% Diabetes mellitus: 17% Oligohydramnios: 10% Hypertensive disorders: 10%	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Admission to NICU 5) Meconium 6) Time to (vaginal) delivery 7) Vaginal deliveries within 24 hours 8) Cost of study medication 9) Time to vaginal delivery 10) Vaginal delivery within 24 hours	1) Apgar score < 7 at 1 minute: Misoprostol: 6/49 (12%) PGE_2: 4/50 (8%) p = 0.525 2) Apgar score < 7 at 5 minutes: Misoprostol: 0/49 PGE_2: 0/50 p = not significant 3) Birthweight (mean ± SD): Misoprostol: 3326.8 ± 529.7 g PGE_2: 3331.4 ± 509.7 g p = 0.965 4) Admission to NICU: Misoprostol: 0/49 PGE_2: 0/50 p = not significant 5) Meconium: Misoprostol: 4/49 (8%) PGE_2: 5/50 (10%) p = 0.950 6) Time to (vaginal) delivery (mean ± SD): Misoprostol (n = 39): 11.4 ± 5.9 hours PGE_2 (n = 40): 18.9 ± 12.7 hours p = 0.001 7) Vaginal deliveries within 24 hours: Misoprostol: 39/39 (100%) PGE_2: 27/40 (68%) p = 0.001 8) Cost of study medication: Misoprostol: $0.20 per dose PGE_2: $65 per kit (no p-value reported) 9) Time to vaginal delivery (mean ± SD): *(continued on next page)*	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (18% of total study population). Sample size estimates based on time to delivery. Study underpowered to detect differences at baseline and for some outcomes – e.g.: 1) Nulliparous with Bishop score ≤ 3: 61% misopostol, 48% PGE_2; p = not significant, but study insufficiently powered. Bias against misoprostol. 2) Prior C-section: 10% misoprostol, 20% PGE_2; bias in favor of misoprostol.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
		Nonreassuring FHR: 8% IUGR: 5% Other: 4%		*Among nulliparas:* Misoprostol (n = 16): 14.4 ± 6.5 hours PGE_2 (n = 16): 26.7 ± 14.3 hours p = 0.004 *Among multiparas:* Misoprostol (n = 23): 9.4 ± 4.7 hours PGE_2 (n = 24): 13.8 ± 8.3 hours p = 0.032 10) Vaginal delivery within 24 hours Misoprostol: 39/39 (100%) PGE_2: 27/40 (68%) p = 0.001	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Crane, Bennett, Young, et al., 1997	Design: RCT, randomization by computer-generated random numbers and sealed envelopes; stratified by status of cervix at initial exam Interventions: 1) Sweeping of membranes (n = 76) Protocol: "As much membrane as possible" separated from lower segment by circumferential sweeping of examining finger two times. Performed between 38 and 40 weeks. "Vigorous" massage by rubbing external os in circular manner if cervix closed. 2) Control exam only (n = 74) Dates: NR Location: Newfoundland, Canada Setting: University hospital (antenatal clinic) Type(s) of providers: NR Length of follow-up: None	No. of subjects at start: 150 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 150 Inclusion criteria: Low-risk pregnancy; gestational age 38-40 weeks Exclusion criteria: Medical disease; pregnancy complications; fetal growth restriction; history of perinatal mortality or low birthweight infant; PROM; abnormal presentation; placenta previa; scheduled cesarean section; other contraindications to vaginal delivery Age (mean ± SD): Sweeping, 27.9 ± 4.8; control, 28.3 ± 4.4 Race: 95% white Gestational age at entry: Sweeping, 39.7 weeks; control, 39.5 weeks Dating criteria: "Firm" LMP or ultrasound prior to 18 weeks Parity: Sweeping: median, 0; 61% nulliparous; control: median, 1.0; 47% nulliparous (p = 0.10) Bishop score: Sweeping: Median, 5; 28% < 7 Control: Median, 5; 16% < 7	1) Spontaneous labor within 7 days 2) Spontaneous labor before 41 weeks 3) Spontaneous labor 4) C-section 5) Epidural 6) PROM 7) Maternal infection 8) Apgar score < 7 at 1 minute 9) Apgar score < 7 at 5 minutes	1) Spontaneous labor within 7 days: Sweeping: 33% Control: 38% p = 0.39 2) Spontaneous labor before 41 weeks: Sweeping: 45% Control: 51% p = 0.66 3) Spontaneous labor: Sweeping: 54% Control: 68% 4) C-section: Sweeping: 13% Control: 14% 5) Epidural: Sweeping: 66% Control: 43% p = 0.006 6) PROM: Sweeping: 6.6% Control: 22% p = 0.008 7) Maternal infection: Sweeping: 6.6% Control: 8.1% 8) Apgar score < 7 at 1 minute: Sweeping: 12% Control: 5.4% p = 1.0 9) Apgar score < 7 at 5 minutes: Sweeping: 0 Control: 0	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: +/- Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + No differences observed when results stratified by open cervix or by parity. More nulliparous women, with less favorable cervix, in sweeping group. Secondary multivariate analyses: Logistic regression: Bishop score < 7, gestational age at entry both predictors of spontaneous labor within 7 days. Log-rank test done for number of days to delivery: median 6.5 for sweeping, 8 for control (p = 0.88). Not clear whether study powered to detect this difference.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Doany and McCarty, 1997	Design: RCT, randomization by table of random numbers Interventions: 1) No membrane stripping + placebo gel (n = 28) Protocol: Placebo gel (4 ml) placed, via syringe, in posterior vaginal fornix. Continuous external fetal and uterine monitoring for 1 hour; if no sign of fetal distress, then patient allowed to go home (instructed to do daily kick counts). Repeat testing at 294 days and every 3-4 days after that. Treatment re-administered at each visit after obtaining reactive NST, normal AFI, and Bishop score. Patients referred to labor and delivery suite if painful contractions every 5 minutes, spontaneous amniorrhexis, decreased fetal movement, nonreactive NST, oligo-hydramnios (AFI < 5), fetal distress, hyperstimulation, or attainment of 307 days of gestation. Labor and delivery managed by appropriate staff (not part of controlled trial). 2) No membrane stripping + PGE_2 gel (n = 37) Protocol: Same as 1), above, except that PGE_2 gel (2 mg) substituted for placebo 3) Membrane stripping + placebo gel (n = 50) Protocol: For membrane stripping, examining finger introduced into the cervical	No. of subjects at start: 150 Dropouts: 7 Loss to follow-up: NA No. of subjects at end: 143 Inclusion criteria: Singleton pregnancy; cephalic presentation; referred for fetal surveillance at ≥ 287 days; reactive NST; AFI 5-25 cm; fetal weight 2500-4500 g; contractions less frequent than every 5 minutes Exclusion criteria: No prenatal care; previous uterine surgery; acute or chronic medical or psychiatric illness; drug use Age (median, with range): No stripping + placebo: 23 (19-26) No stripping + PGE_2: 23 (21-30) Stripping + placebo: 22 (19-26) Stripping + PGE_2: 25 (22-27) Race: No stripping + placebo: 100% Hispanic No stripping + PGE_2: 100% Hispanic Stripping + placebo: 94% Hispanic Stripping + PGE_2: 96% Hispanic Gestational age at entry (median, with 25-75th percentile): No stripping + placebo: 288 days (287-290) No stripping + PGE_2: 288 days (287-291) Stripping + placebo: 288 days (287-290)	1) Apgar score < 7 at 5 minutes 2) Birthweight 3) Admission to NICU 4) Probable neonatal sepsis 5) Amnionitis 6) Preeclampsia 7) Maternal hemorrhage 8) Gestational age at delivery 9) Inductions 10) Oxytocin augmentation 11) Meconium 12) C-sections 13) Operative vaginal deliveries 14) Time from enrollment to delivery	1) Apgar score < 7 at 5 minutes: No stripping + placebo: 0 No stripping + PGE_2: 3% Stripping + placebo: 4% Stripping + PGE_2: 4% p = 0.99 2) Birthweight (mean [in grams] ± SD): No stripping + placebo: 3613 ± 273 No stripping + PGE_2: 3527 ± 333 Stripping + placebo: 3605 ± 365 Stripping + PGE_2: 3614 ± 479 p = 0.70 3) Admission to NICU: No stripping + placebo: 0 No stripping + PGE_2: 5% Stripping + placebo: 2% Stripping + PGE_2: 4% p = 0.70 4) Probable neonatal sepsis: No stripping + placebo: 7% No stripping + PGE_2: 11% Stripping + placebo: 6% Stripping + PGE_2: 7% p = 0.86 5) Amnionitis: No stripping + placebo: 0 No stripping + PGE_2: 11% Stripping + placebo: 10% Stripping + PGE_2: 11% p = 0.32 6) Preeclampsia: No stripping + placebo: 0 No stripping + PGE_2: 14% Stripping + placebo: 0 Stripping + PGE_2: 7% p = 0.01	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Results not stratified by parity. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	canal and a total of 3 circumferential sweeps made between the lower uterine segment and the chorionic membranes. When cervical canal not accessible, then cervix pulled anteriorly and massaged. Rest of protocol as in 1), above. 4) Membrane stripping + PGE_2 gel (n = 28) Protocol: Membrane stripping as in 3), above. Rest of protocol as in 2), above. Dates: NR Location: Sylmar, CA Setting: University hospital Type(s) of providers: Not specified Length of follow-up: None	Stripping + PGE_2: 288 days (287-289) Dating criteria: LMP confirmed by uterine size, fetal heart tones, and U/S (no date given) Parity (% nulliparous): No stripping + placebo: 54% No stripping + PGE_2: 38% Stripping + placebo: 50% Stripping + PGE_2: 43% Bishop score (% ≤ 6): No stripping + placebo: 50% No stripping + PGE_2: 69% Stripping + placebo: 63% Stripping + PGE_2: 63%		7) Maternal hemorrhage: No stripping + placebo: 7% No stripping + PGE_2: 0 Stripping + placebo: 0 Stripping + PGE_2: 4% p = 0.05 8) Gestational age at delivery (median [in days], with 25-75th percentile): No stripping + placebo: 297 (292-302) No stripping + PGE_2: 294 (290-298) Stripping + placebo: 294 (291-298 Stripping + PGE_2: 290 (289-293) p = 0.005 9) Inductions: No stripping + placebo: 33% No stripping + PGE_2: 28% Stripping + placebo: 27% Stripping + PGE_2: 14% p = 0.42 10) Oxytocin augmentation: No stripping + placebo: 48% No stripping + PGE_2: 47% Stripping + placebo: 37% Stripping + PGE_2: 36% p = 0.65 11) Meconium: No stripping + placebo: 30% No stripping + PGE_2: 19% Stripping + placebo: 26% Stripping + PGE_2: 21% p = 0.67 12) C-sections: No stripping + placebo: 4% No stripping + PGE_2: 8% Stripping + placebo: 8% Stripping + PGE_2: 11% p = 0.08	

(continued on next page)

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				13) Operative vaginal deliveries: No stripping + placebo: 4% No stripping + PGE_2: 3% Stripping + placebo: 18% Stripping + PGE_2: 7% (no p-value reported) 14) Time from enrollment to delivery (median [in days], with 25-75th percentile): No stripping + placebo: 7 (3.5-11.5) No stripping + PGE_2: 2 (0-7) Stripping + placebo: 4 (2-8) Stripping + PGE_2: 1 (0-4) p = 0.001	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Elliott, Brennand, and Calder, 1998	Design: RCT, randomization method not detailed but implied by computer-generated random numbers Interventions: 1) Mifepristone 50 mg (n = 25) Protocol: 50 mg given orally in women with indication for induction between 37 weeks and 41 weeks, 4 days. 2) Mifepristone 200 mg (n = 25) Protocol: Same as above, except dose 200 mg. 3) Placebo (n = 30) In all groups, patients had NST and cervical exam at 24 and 48 hours after initial dose. Induction scheduled for 72 hours after medication if no labor. Induction performed using 1 mg PGE_2 gel as initial dose, with oxytocin as clinically indicated. Dates: NR Location: Edinburgh, UK Setting: University hospital Type(s) of providers: NR Length of follow-up: None	No. of subjects at start: 80 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 80 Inclusion criteria: Single gestation; vertex presentation; Bishop score ≤ 4 Exclusion criteria: Signs and symptoms of labor; placental insufficiency; contraindications to mifepristone Age (mean ± SD): Placebo: 26.2 ± 5.9 50 mg: 25.8 ± 4.5 200 mg:25.6 ± 3.3 Race: NR Gestational age at entry (mean ± SD): Placebo: 40 weeks, 6 days (± 3.6 days) 50 mg: 40 weeks, 5 days (± 5.5 days) 200 mg: 40 weeks, 6 days (± 5.1 days Dating criteria: 1st trimester U/S Parity: 100% nulliparous Bishop score (median, with range): Placebo: 3 (1-4) 50 mg: 4 (2-4) 200 mg: 3 (1-4)	1) Proportion in spontaneous labor within 72 hours 2) Proportion with Bishop score ≥ 6 at induction 3) Time to onset of labor 4) Time to delivery 5) Fetal distress in labor requiring intervention 6) Cesarean delivery 7) Neonatal outcomes	1) Proportion in spontaneous labor within 72 hours: Placebo: 23.3% 50 mg: 32% 200 mg: 36% 2) Proportion with Bishop score ≥ 6 at induction: Placebo: 6.7% 50 mg: 16% 200 mg: 28% 3) Time to onset of labor (median): Placebo: 81 hours 15 minutes 50 mg: 80 hours 20 minutes 200 mg: 75 hours 50 minutes 4) Time to delivery (median): Placebo: 88 hours 14 minutes 50 mg: 85 hours 15 minutes 200 mg: 84 hours 6 minutes 5) Fetal distress in labor requiring intervention: Placebo: 13.3% 50 mg: 24% 200 mg: 48% 6) Cesarean delivery: Placebo: 25% 50 mg: 5% 200 mg: 38% p=0.033, Placebo vs. 50 mg p=0.075, Placebo vs. 200 mg 200 mg group: 8/9 for fetal distress, 1 for dystocia Placebo: 3/8 for fetal distress, 5 for dystocia 7) Neonatal outcomes: Jaundice: Placebo: 6.7% 50 mg: 8%	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Study underpowered to detect differences in cesarean rates, neonatal outcomes. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				200 mg: 28% Trends toward lower ACTH, higher cortisol in infants in 200 mg group	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Elliott and Flaherty, 1984	Design: RCT, randomization by table of random numbers Interventions: 1) Breast stimulation (n = 100) Protocol: Patients instructed to manually stimulate the nipple, areola, and distal breast with the balls of the fingertips, one breast at a time, for 15 minutes at a time, for 1 hour. Encouraged to do this 3 x per day (total of 3 hours per day). Re-evaluation at 42 weeks. If Bishop score ≥ 8, then labor induced. If Bishop score < 8, then CST administered. If CST reactive (negative), then further week of treatment. If CST abnormal, then labor induced. 2) Pelvic exam (control) (n = 100) Protocol: Pelvic exam given. Patients instructed to abstain from sexual intercourse and to avoid breast stimulation. Re-evaluation at 42 weeks. If Bishop score ≥ 8, then labor induced. If Bishop score < 8, then CST administered. If CST abnormal, then labor induced. If CST reactive (negative), then patient randomly assigned a second time to breast stimulation or control for further treatment. Dates: NR Location: San Francisco, CA	No. of subjects at start: 200 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 200 Inclusion criteria: Uncomplicated prenatal course; ≥ 39 weeks gestation Exclusion criteria: None specified Age (mean ± SD): Breast stimulation, 25.0 ± 4.75; control, 24.4 ± 4.88 Race: NR Gestational age at entry: NR; all subjects "approximately" 39 weeks Dating criteria: Reliable menstrual history, early pregnancy test, early vaginal estimation of uterine size, fetal heart auscultation at 20 weeks, and/or obstetric sonograms Parity (mean ± SD): Breast stimulation, 0.79 ± 1.04; control, 0.84 ± 1.10 Bishop score (mean ± SD): Breast stimulation, 4.67 ± 2.27; control, 4.15 ± 2.34	1) Apgar scores < 7 at 1 minute 2) Apgar scores < 7 at 5 minutes 3) Birthweight 4) Meconium aspiration 5) Meconium in labor 6) Inductions 7) C-sections 8) Dysmature infant 9) Death 10) Proportion of patients reaching 43 weeks with Bishop score < 8	1) Apgar scores < 7 at 1 minute: *Among women delivering at ≤ 42 weeks:* Breast stimulation: 6/95 (6%) Control: 1/83 (1%) p = not significant *Among women delivering at > 42 weeks:* Breast stimulation: 1/5 (20%) Control: 2/17 (12%) p = not significant 2) Apgar scores < 7 at 5 minutes: *Among women delivering at ≤ 42 weeks:* Breast stimulation: 1/95 (1%) Control: 0 p = not significant *Among women delivering at > 42 weeks:* Breast stimulation: 0 Control: 0 3) Birthweight (mean ± SD): Breast stimulation: 3594 ± 441 g Control: 3649 ± 394 g p = not significant 4) Meconium aspiration: Breast stimulation: 0 Control: 0 5) Meconium in labor: *Among women delivering at ≤ 42 weeks:* Breast stimulation: 25/95 (26%) Control: 22/83 (26%) p = not significant *Among women delivering at > 42 weeks:* Breast stimulation: 0 Control: 11/17 (65%) p < 0.01	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: +

(continued on next page)

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Setting: Military hospital Type(s) of providers: General OB/GYN Length of follow-up: None			6) Inductions: *Among women delivering at ≤ 42 weeks:* Breast stimulation: 6/95 (6%) Control: 6/83 (7%) p = not significant *Among women delivering at > 42 weeks:* Breast stimulation: 0 Control: 2/17 (12%) p = not significant 7) C-sections: *Among women delivering at ≤ 42 weeks:* Breast stimulation: 9/95 (9%) Control: 5/83 (6%) p = not significant *Among women delivering at > 42 weeks:* Breast stimulation: 0 Control: 5/17 (29%) p = not significant 8) Dysmature infant: Breast stimulation: 3/100 (3%) Control: 5/100 (5%) p = not significant 9) Death: Breast stimulation: 0/100 Control: 0/100 p = not significant 10) Proportion of patients reaching 43 weeks with Bishop score < 8: Breast stimulation: 5/100 Control: 17/100 p < 0.01	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Escudero and Contreras, 1997	Design: RCT, randomization by table of random numbers Interventions: 1) Misoprostol (n = 53) Protocol: Misoprostol 50 µg placed in posterior vaginal fornix. Dose repeated every 4 hours until adequate labor achieved (≥ 3 contractions of 40-50 seconds each in 10 min). Maximum total dose 350 µg. AROM performed as soon as possible. Patients with arrest of dilatation managed with oxytocin infusion, as below. 2) Oxytocin (n = 67) Protocol: Oxytocin infusion started at 4 mIU/min for 45 minutes, then increased by 2 mIU/min at 15-minute intervals up to 20 mIU/min. Dates: Sep 1994 - Mar 1995 Location: Lima, Peru Setting: University hospital Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: 123 Dropouts: 3 (excluded from analysis due to protocol violations) Loss to follow-up: NA No. of subjects at end: 120 Inclusion criteria: Obstetric or medical indication for induction; no labor or fetal distress; no previous uterine; singleton pregnancy with vertex presentation; no contraindication to vaginal delivery Exclusion criteria: None specified Age (mean ± SD): Misoprostol, 27.1 ± 6.1; oxytocin, 25.5 ± 6.0 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 39.0 ± 2.2 weeks; oxytocin, 39.3 ± 2.1 weeks Dating criteria: NR Parity (mean ± SD): Misoprostol, 0.8 ± 1.2; oxytocin, 0.5 ± 1.0 Bishop score (mean ± SD): Misoprostol, 2.6 ± 1.5; oxytocin, 2.9 ± 1.5 Other: Indications for induction: Preeclampsia: 43% Postterm: 25% PROM: 25% Fetal demise: 4% Other: 3%	1) Apgar scores at 1 minute 2) Apgar scores at 5 minutes 3) Birthweight 4) Interval from induction to delivery 5) C-sections 6) Vaginal deliveries within 24 hours 7) Hyperstimulation 8) Any labor complication	1) Apgar scores at 1 minute (mean ± SD): Misoprostol (n = 51): 8.0 ± 1.4 Oxytocin (n = 41): 8.0 ± 1.5 p = 1.0000 2) Apgar scores at 5 minute (mean ± SD): Misoprostol (n = 51): 9.1 ± 0.9 Oxytocin (n = 41): 9.0 ± 1.3 p = 0.6646 3) Birthweight (mean ± SD): Misoprostol (n = 55): 3090.5 ± 556.9 g Oxytocin (n = 41): 3254.4 ± 493.2 g p = 0.1378 4) Interval from induction to delivery (mean ± SD): Misoprostol: 11.3 ± 6.9 hours Oxytocin: 8.4 ± 4.1 hours p = 0.0050 5) C-sections: Misoprostol : 10/57 (17.6%) Oxytocin: 4/63 (6.4%) p = 0.0560 6) Vaginal deliveries within 24 hours: Misoprostol: 45/57 (78.9%) Oxytocin: 37/63 (58.7%) p = 0.0017 7) Hyperstimulation: Misoprostol: 5/57 (8.8%) Oxytocin: 0/63 p = 0.0160 8) Any labor complication: Misoprostol: 12/57 (21.1%) Oxytocin: 5/63 (7.9%) p = 0.0400	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: - Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (25% of total study population). Results not stratified by parity.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Evans, Dougan, Moawad, et al., 1983	Design: RCT, method of randomization not described Interventions: 1) Relaxin 4 mg (n = 10) Protocol: 4-mg pellet inserted into, or placed closely against, the cervix, as permitted by cervical dilatation. Cervical diaphragm placed behind the pellet to maintain its position until it dissolved (approxi-mately 30 minutes). Patient then allowed to go home. Standard management protocol of estriols 3 times per week and NSTs 1-2 times per week was followed. If patient reached 42 weeks' gestation, then she was admitted for induction. 2) Relaxin 2 mg (n = 13) Protocol: Same as above, except that 2-mg pellet used. 3) Placebo (n = 14) Protocol: Same as above, except that placebo pellet used. Dates: NR Location: Chicago, IL Setting: University hospital Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: 37 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 37 Inclusion criteria: ≥ 41 weeks gestation; scheduled to undergo oxytocin induction of labor Exclusion criteria: None specified Age (mean ± SD): Relaxin 4 mg: 26.0 ± 5.7 Relaxin 2 mg: 23.3 ± 5.4 Placebo: 21.3 ± 4.4 Race: NR Gestational age at entry (mean ± SD): Relaxin 4 mg: 41.0 ± 0.2 weeks Relaxin 2 mg: 41.2 ± 0.3 weeks Placebo: 41.1 ± 0.2 weeks Dating criteria: NR Parity (mean ± SD): Relaxin 4 mg: 1.0 ± 1.2 Relaxin 2 mg: 1.2 ± 1.1 Placebo: 1.1 ± 0.9 Bishop score: NR Other: Initial cervical coefficient (dilatation x % effacement): Relaxin 4 mg: 38.0 ± 44.5 Relaxin 2 mg: 49.6 ± 44.4 Placebo: 70.0 ± 62.6	1) Apgar scores at 5 minutes 2) Birthweight 3) Days to admission 4) Number admitted in labor 5) Time to delivery	1) Apgar scores at 5 minutes (mean ± SD): Relaxin 4 mg: 8.6 ± 1.2 Relaxin 2 mg: 9.0 ± 0.4 Placebo: 9.0 ± 0.4 p = not significant 2) Birthweight (mean ± SD): Relaxin 4 mg: 3113 ± 447 g Relaxin 2 mg: 3256 ± 613 g Placebo: 3245 ± 479 g p = not significant 3) Days to admission (mean ± SD): Relaxin 4 mg: 4.6 ± 1.6 Relaxin 2 mg: 5.3 ± 2.2 Placebo: 5.3 ± 2.1 p = not significant 4) Number admitted in labor: Relaxin 4 mg: 3/10 (30%) Relaxin 2 mg: 7/13 (54%) Placebo: 6/14 (43%) p = not significant 5) Time to delivery (mean ± SD): Relaxin 4 mg: 11.3 ± 7.2 hours Relaxin 2 mg: 7.7 ± 5.0 hours Placebo: 14.8 ± 12.2 hours p = not significant	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Article describes two trials; only the trial conducted on "postdate" women abstracted here. Investigators used the "cervical coefficient" (dilatation x % effacement) instead of the Bishop score as a measure of cervical ripeness. See Hendricks CH, Brenner WE, Kraus G. Normal cervical dilatation pattern in late pregnancy and labor. Am J Obstet Gynecol 1970;106: 1065-82. Improvement in time to delivery in both nullipara and multipara.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Farah, Sanchez-Ramos, Rosa, et al., 1997	Design: RCT, randomization by computer-generated table of random numbers Interventions: 1) Misoprostol 25 µg (n = 192) Protocol: Tablet placed in posterior vaginal fornix. Dose repeated every 3 hours until adequate labor achieved (≥ 3 contractions/10 minutes). Maximum total dose 200 µg, or 8 applications. 2) Misoprostol 50 µg (n = 207) Protocol: Same as above, except maximum total dose 400 µg. In both groups, amniotomy performed as soon as cervical dilation permitted. Patients in active phase of labor with arrest of dilation and those who failed to achieve active labor after the maximum dose of misoprostol were given oxytocin. Dates: July 1994 - Sep 1995 Location: Jacksonville and Gainesville, FL Setting: 2 university hospitals Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: 430 Dropouts: 31 Loss to follow-up: NA No. of subjects at end: 399 Inclusion criteria: Obstetric or medical indication for induction; Bishop score < 5; no active labor or fetal distress; no history of uterine surgery; singleton 3rd-trimester pregnancy; vertex presentation; no contraindication to vaginal delivery; no contra-indication to prostaglandins Exclusion criteria: None specified Age (mean ± SD): 25 µg, 23.8 ± 6.2; 50 µg, 23.7 ± 6.4 Race: 25 µg, 52% non-White; 50 µg, 59% non-White Gestational age at entry (mean ± SD): 25 µg, 28.9 ± 2.3 weeks; 50 µg, 38.4 ± 2.8 weeks Dating criteria: NR Parity: 25 µg, 59% nulliparous; 50 µg, 60% nulliparous Bishop score: 25 µg, 86% < 6; 50 µg, 88% < 6 Other: Indications for induction: PROM: 27% Pregnancy-induced hypertension: 22% Postterm: 14% IUGR: 8%	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Cord pH < 7.6 4) Mean cord pH 5) Admission to NICU 6) Interval from induction to delivery 7) C-sections 8) Tachysystole 9) Hyperstimulation 10) Delivery within 24 hours	1) Apgar score < 7 at 1 minute: 25-µg dose: 33/192 (17.2%) 50-µg dose: 39/207 (18.8%) p = not significant 2) Apgar score < 7 at 5 minutes: 25-µg dose: 1/192 (0.5%) 50-µg dose: 7/207 (3.4%) p = 0.07 3) Cord pH < 7.6: 25-µg dose: 13/192 (6.8%) 50-µg dose: 27/207 (13.0%) p = 0.04 4) Mean cord pH (± SD): 25-µg dose: 7.26 ± 0.07 50-µg dose: 7.25 ± 0.09 p = not significant 5) Admission to NICU: 25-µg dose: 11/192 (5.7%) 50-µg dose: 23/207 (11.1%) p = 0.07 6) Interval from induction to delivery (mean ± SD): 25-µg dose: 970 ± 684 minutes 50-µg dose: 826 ± 554 minutes p = 0.02 7) C-sections: 25-µg dose: 23/192 (12%) 50-µg dose: 33/207 (15.9%) p = not significant 8) Tachysystole: 25-µg dose: 30/192 (15.6%) 50-µg dose: 68/207 (32.8%) p = 0.0001 9) Hyperstimulation: 25-µg dose: 10/192 (5.2%) 50-µg dose: 12/207 (5.8%)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (14% of total study population). Sample size estimates based on incidence of tachysystole. Differences in indications for C-sections (e.g., fetal distress 30% 25 µg vs. 48.5% 50 µg; difference not significant, but study underpowered). *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
		Abnormal FHR: 5% Diabetes mellitus: 3% Other: 21%		p = not significant 10) Delivery within 24 hours: 25-µg dose: 79/192 (41.1%) 50-µg dose: 101/207 (48.8%) p = not significant	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Fletcher, Mitchell, Frederick, et al., 1994	Design: RCT, randomization by drawing odd/even numbers in sealed envelopes Interventions: 1) Misoprostol (n = 32) Protocol: 100-µg tablet placed in posterior vaginal fornix. 2) PGE_2 (n = 31) Protocol: 3-mg tablet placed in posterior vaginal fornix. In both groups, patients not in labor at 12 hours were sent to the labor ward for oxytocin infusion. Dates: Sep-Oct 1992 Location: Kingston, Jamaica Setting: University hospital Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: 64 Dropouts: 1 (excluded from analysis due to protocol violation) Loss to follow-up: NA No. of subjects at end: 63 Inclusion criteria: Scheduled for induction Exclusion criteria: Known contraindications to vaginal prostaglandins, including a previous scar on the uterus; antepartum hemorrhage; fetal distress; PROM; abnormal lie; cephalopelvic disproportion; any maternal illness for which induction contraindicated Age (mean ± SD): Misoprostol, 27.1 ± 6.0; PGE_2, 28.0 ± 5.1 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 38.8 ± 2.8 weeks; PGE_2, 39.7 ± 1.5 weeks Dating criteria: NR Parity (mean ± SD): Misoprostol, 0.6 ± 0.8; PGE_2, 1.1 ± 1.2 Bishop score (mean ± SD): Misoprostol, 4.1 ± 2.3; 4.4 ± 2.5 Other: Indications for induction: Hypertension: 38% Postterm: 33% Diabetes: 11%	1) Apgar scores at 1 and 5 minutes 2) Perinatal deaths 3) Time from induction to delivery 4) Forceps deliveries 5) Vacuum deliveries 6) C-sections	1) Apgar scores at 1 and 5 minutes (mean): *At 1 minute:* Misoprostol: 7.6 PGE_2: 8.3 p = 0.12 *At 5 minutes:* Misoprostol: 8.8 PGE_2: 9.1 p = 0.45 2) Perinatal deaths: None in either group 3) Time from induction to delivery (mean ± SD): Misoprostol: 21.8 ± 29.3 hours PGE_2: 32.3 ± 36.6 hours p = 0.21 4) Forceps deliveries: Misoprostol: 1/32 (3%) PGE_2: 0/31 (no p-value reported) 5) Vacuum deliveries: Misoprostol: 3/32 (9%) PGE_2: 0/32 (no p-value reported) 6) C-sections: Misoprostol: 1/32 (3%) PGE_2: 3/31 (10%) p = 0.17	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: - Interventions described: ?? Mode of delivery: ?? Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (33% of total study population). Results not stratified by parity. Study underpowered to detect differences. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
		Excess weight gain: 3% Cardiac: 3% IUGR, previous stillbirth, poor weight gain, eclampsia, low biological profile score, weight loss at term, and unstable lie: 1.6% each			

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Fletcher, Mitchell, Simeon, et al., 1993	Design: RCT, method of randomization not described Interventions: 1) Misoprostol (n = 24) Protocol: Misoprostol 100 µg powder mixed with sterile gel and placed in posterior vaginal fornix using a syringe. At 12 hours, patients not in labor were sent to the labor ward for oxytocin infusion. 2) Placebo (n = 21) Protocol: Same as above, except placebo powder (0.05 mg ethinyl oestradiol) used instead of misoprostol. Dates: NR Location: Kingston, Jamaica Setting: University hospital Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: 48 Dropouts: 3 Loss to follow-up: NA No. of subjects at end: 45 Inclusion criteria: Indication for induction; 3rd trimester pregnancy; unripe cervix; no contraindication to prostaglandins Exclusion criteria: None specified Age (mean ± SD): Misoprostol, 25.8 ± 6.3; placebo, 26.0 ± 4.9 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 39.5 ± 2.2 weeks; placebo, 39.8 ± 1.7 weeks Dating criteria: NR Parity: Misoprostol, 54% nulliparous; placebo, 43% nulliparous Bishop score (mean ± SD): Misoprostol, 3.1 ± 1.5; placebo, 3.1 ± 2.0 Other: Indications for induction: Postterm: 51% Preeclampsia: 27% Preeclampsia with IUD: 4% Diabetes mellitus: 7% IUGR: 2% UTI: 2% Rheumatic heart: 2% Previous stillbirth: 2% Oligohydramnios: 2%	1) Apgar scores at 1 and 5 minutes (for women receiving oxytocin augmentation) 2) Meconium staining 3) Fetal tachycardia 4) Time from induction to delivery 5) Forceps deliveries 6) C-sections	1) Apgar scores at 1 and 5 minutes (mean ± SD) (for women receiving oxytocin augmentation): *At 1 minute:* Misoprostol (n = 7): 8.1 ± 2.3 Placebo (n = 13): 7.7 ± 2.2 p = 0.34 *At 5 minutes:* Misoprostol (n = 7): 8.9 ± 2.2 Placebo (n = 13): 8.9 ± 2.2 p = 0.73 2) Meconium staining: Misoprostol: 2/24 (8%) Placebo: 0/21 p = not significant 3) Fetal tachycardia: Misoprostol: 0/24 Placebo: 2/21 (9.5%) p = not significant 4) Time from induction to delivery (mean ± SD): Misoprostol: 15.6 ± 12.5 hours Placebo: 43.2 ± 20.5 hours p < 0.001 5) Forceps deliveries: Misoprostol: 1/24 (4%) Placebo: 1/21 (5%) p = not significant 6) C-sections: Misoprostol: 2/24 (8%) Placebo: 3/21 (14%) p = not significant	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (51% of total study population). Results not stratified by parity.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Frydman, Lelaidier, Baton-Saint-Mleux, et al., 1992	Design: RCT, randomized by computer-generated tables Interventions: 1) Mifepristone (n = 60), in women from 37.5-41.4 weeks, given as two 200-mg oral doses 24 hours apart 2) Placebo (n = 60) In both groups, NST performed each day until day 4, when induction done with vaginal PGE_2 if no labor. Dates: Apr 1990 - Jan 1991 Location: Clamart, France Setting: Unspecified hospital Type(s) of providers: NR Length of follow-up: None	No of subjects at start: 120 Drop-outs: 8 Loss to follow-up: NA No of subjects at end: 112 Inclusion criteria: Indication for induction (48% "postdates"); Bishop score < 4 Exclusion criteria: Medical condition; nonvertex presentation; more than one prior cesarean; multiple gestation; premature rupture of membranes Age (mean ± SD) Mifepristone: 31 ± 4.1 Placebo: 29 ± 3.6 Gestational age at entry (mean ± SD): Mifepristone: 39.9 ± 1.2 Placebo: 39.7 ± 1.2 Parity (% nulliparous) Mifepristone: 65% Placebo: 60% Bishop score: NR (100% < 4)	1) Proportion in spontaneous labor 2) Bishop score < 4 on day 4 3) Interval from randomization to start of labor 4) Cesarean delivery 5) Epidural anesthesia 6) Apgar < 7 at 1 minute 7) Apgar <7 at 5 minutes	1) Proportion in spontaneous labor: Mifepristone: 54% Placebo: 18% p < 0.001 2) Bishop score < 4 on day 4: Mifepristone: 23% Placebo: 58% p < 0.001 3) Interval from randomization to start of labor: Mifepristone: mean 51 h 45 min Placebo: mean 74 h 30 min P < 0.001 4) Cesarean delivery: Mifepristone: 30% Placebo: 30% No detectable differences by indication 5) Epidural anesthesia: Mifepristone: 73% Placebo: 82% p = not significant 6) Apgar < 7 at 1 minute: Mifepristone: 5/57 Placebo: 4/55 p = not significant NS 7) Apgar <7 at 5 minutes: Mifepristone: 0/57 Placebo: 0/55	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: - Mode of delivery: + Sample size: - Statistical tests: - Gestational age: + Dating criteria: + Bishop score: - Study underpowered to detect differences in categorical outcomes.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Garry, Figueroa, Guillaume, et al., 2000	Design: RCT, patients alternately assigned to one of two study groups Interventions: 1) Castor oil (n = 52) Protocol: Single 60-ml oral dose given, diluted in apple or orange juice. 2) No treatment (n = 48) Dates: July 1992 - Feb 1993 Location: Brooklyn, NY Setting: Community hospital Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: 103 Dropouts: 3 Loss to follow-up: NA No. of subjects at end: 100 Inclusion criteria: Gestational age 40-42 weeks; Bishop score ≤ 4; no regular uterine contractions Exclusion criteria: Ruptured membranes; multiple gestations; oligohydramnios; IUGR; abnormal FHR tracings; biophysical profile score ≤ 8; noncephalic presentation; maternal medical complications Age (mean ± SD): Castor oil, 24.8 ± 6.7; no treatment, 24.4 ± 4.9 Race: NR Gestational age at entry (mean ± SD): Castor oil, 284.4 ± 4.2 days; no treatment, 284.7 ± 3.6 days Dating criteria: LMP or early U/S (obtained in 1st or 2nd trimester) Parity: Castor oil, 42.3% nulliparous; no treatment, 43.8% nulliparous Bishop score: NR; score ≤ 4 required for entry into study Other: Indications for induction not reported	1) Birthweight 2) Meconium staining 3) Labor within 24 hours 4) C-sections	1) Birthweight (mean ± SD): Castor oil: 3486 ± 434 g No treatment: 3437 ± 420 g p = 0.56 2) Meconium staining: Castor oil: 10.4% No treatment: 11.5% p = 3) Labor within 24 hours: Castor oil: 30/52 (57.7%) No treatment: 2/48 (4.2%) p < 0.001 4) C-sections: Castor oil: 10/52 (19.2%) No treatment: 4/48 (8.3%) p = 0.20	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: - Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Results not stratified by parity or by indication for induction. Study underpowered to detect differences in C-section rate.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Giacalone, Targosz, Laffargue, et al., 1998	Design: RCT, randomization by permutation blocks and sealed envelope Interventions: 1) Mifepristone for cervical ripening (n = 41) Protocol: Mifepristone 400 mg given as a single oral dose. Patients re-examined 24 and 48 hours later. If Bishop score ≥ 6, then patient induced with oxytocin and amniotomy. If Bishop score < 6, then cervical ripening/induction considered to have failed, and patient managed in accordance with physician's "usual induction techniques." FHR tracing done at each exam visit and during labor. 2) Placebo (n = 42) Protocol: Same as above, but with identical placebo used in place of mifepristone. Dates: Jan 1991 - Feb 1992 Location: Montpellier and Nantes, Frances Setting: 2 university hospitals Type(s) of providers: Unspecified OB/GYN Length of follow-up: Follow-up visit scheduled for neonates 1-2 months after birth	No. of subjects at start: 84 Dropouts: 1 Loss to follow-up: 7 (not available for 1-2 month follow-up) No. of subjects at end: 76 Inclusion criteria: Gestational age ≥ 41 weeks and 3 days; Bishop score < 6; labor induction post-ponable for 48 hours Exclusion criteria: Contra-indication to vaginal delivery; multiple gestation; > 4 previous deliveries; uterine scar; premature rupture of the membranes; FHR abnormality; impaired renal, adrenal, or hepatic function; corticosteroid therapy during pregnancy; abnormal hemostasis; anticoagulant therapy Age (mean ± SD): Mifepristone, 28.5 ± 4.3; placebo, 28.3 ± 5.0 Race: NR Gestational age at entry: NR*; at delivery* mifepristone, 41.5 ± 0.2 weeks; placebo, 41.6 ± 0.2 weeks Dating criteria: NR Parity: Mifepristone, 20/41 (49%) nulliparous; placebo, 20/42 (48%) nulliparous Bishop score (median, with range): Mifepristone, 3 (1 to 5); placebo, 3 (1 to 5)	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Umbilical artery pH < 7.2 5) Glycemia ≤ 40 mg/dL 6) Cortisol levels 7) Post-natal abnormalities 8) C-sections 9) Cervical ripening in patients with Bishop score < 6 10) Instrumental delivery 11) Time to onset of labor 12) Time to delivery (excluding C-sections)	1) Apgar score < 7 at 1 minute: Mifepristone: 3/41 (7.3%) Placebo: 2/42 (4.8%) p = not significant 2) Apgar score < 7 at 5 minutes: Mifepristone: 0 Placebo: 0 3) Birthweight (mean ± SD): Mifepristone: 3418 ± 380 g Placebo: 3502 ± 364 g p = not significant 4) Umbilical artery pH < 7.2: Mifepristone: 3/41 (7.3%) Placebo: 2/42 (4.8%) p = not significant 5) Glycemia ≤ 40 mg/dL: *Day 1:* Mifepristone: 1/41 (2.4%) Placebo: 6/42 (14.3%) p = not significant *Day 2:* Mifepristone: 1/41 (2.4%) Placebo: 1/42 (2.4%) p = not significant 6) Cortisol levels (median, with range): Mifepristone: 153.5 nmol/L (42 to 537) Placebo: 94.5 nmol/L (28 to 223) (no p-value reported) 7) Post-natal abnormalities (at 1-2 month follow-up): Mifepristone: 5/38 (13%) Placebo: 2/38 (5.3%) p = 0.42	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not stratified by parity.

(continued on next page)

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				8) C-sections: Mifepristone: 7/41 (17%) Placebo: 6/42 (14.3%) p = not significant 9) Cervical ripening in patients with Bishop score < 6: Mifepristone: 7/41 (17.1%) Placebo: 17/42 (40.4%) p = not significant 10) Instrumental delivery: Mifepristone: 9/41 (22%) Placebo: 6/42 (14.3%) (no p-value reported) 11) Time to onset of labor (median, with range): Mifepristone: 31.7 hours (9.5 to 117.8) Placebo: 53.9 hours (2.5 to 192.0) p = 0.02 12) Time to delivery (excluding C-sections) (median, with range): Mifepristone: 31.3 hours (13.2 to 123.3) Placebo: 58.5 hours (5.8 to 193.7) p = 0.02	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Gottschall, Borgida, Mihalek, et al., 1997	Design: RCT, randomization by random-numbers table and sealed envelopes Interventions: 1) Misoprostol (n = 38) Protocol: 100 µg placed in posterior vaginal fornix. 2) PGE_2 gel (n = 37) Protocol: PGE_2 gel (5 mg) placed in posterior vaginal fornix by syringe. In both groups, patients were re-examined at 6 hours after placement of study medication. If patient in labor (≥ 3 contractions/10 minutes, with changes in cervical dilatation), then amniotomy performed. If patient not in labor, then oxytocin augmentation initiated. Dates: Nov 1995- Aug 1996 Location: New Britain, CT Setting: Community hospital Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: 75 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 75 Inclusion criteria: Indication for cervical ripening and induction; live, singleton fetus; cephalic presentation; intact membranes; reactive FHR tracing; no contra-indications to a vaginal delivery Exclusion criteria: Previous uterine scar; allergy to prostaglandin agents Age (mean ± SD): Misoprostol, 28.4 ± 5.7; PGE_2, 26.9 ± 6.4 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 39.8 ± 1.7 weeks; PGE_2, 39.8 ± 2.2 weeks Dating criteria: NR Parity: Misoprostol, 61% nulli-parous; PGE_2, 68% nulliparous Bishop score (median): 4, both groups Other: Indications for induction: Postterm: 40% Preeclampsia: 27% Oligohydramnios: 16% IUGR: 7% Chronic hypertension: 3% Diabetes: 1% Other: 7%	1) Apgar scores at 1 and 5 minutes 2) Birthweight 3) Time from induction to delivery 4) Delivery by 24 hours 5) Hyperstimulation 6) Tachysystole 7) C-sections	1) Apgar scores at 1 and 5 minutes (median): *At 1 minute:* Misoprostol: 8 PGE_2: 8 p = not significant *At 5 minutes:* Misoprostol: 9 PGE_2: 9 p = not significant 2) Birthweight (mean ± SD): Misoprostol: 3438 ± 536 g PGE_2: 3435 ± 591 g p = not significant 3) Time from induction to delivery (mean ± SD): Misoprostol: 14.7 ± 6.4 hours PGE_2: 20.4 ± 10.2 hours p = 0.005 4) Delivery by 24 hours: Misoprostol: 95% PGE_2: 70% p = 0.005 5) Hyperstimulation: Misoprostol: 2.8% PGE_2: 0 p = not significant 6) Tachysystole: Misoprostol: 15.8% PGE_2: 2.7% p = not significant 7) C-sections: Misoprostol: 18% PGE_2: 27% p = not significant	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (40% of total study population). Sample size estimates based on time to delivery. Underpowered to detect differences in some outcomes. Findings similar when nulliparas analyzed separately.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Graves, Baskett, Gray, et al., 1985	Design: RCT, randomization method not specified Interventions: 1) 3 mg PGE_2 gel (n = 20) inserted into posterior vaginal fornix via catheter 2) 2 mg PGE_2 gel (n = 20) 3) 1 mg PGE_2 gel (n = 20) 4) Placebo gel (n = 20) In all groups, patients monitored for 1 hour after insertion. If no labor after 12-16 hours, induction with oxytocin ± amniotomy. Dates: NR Location: Halifax, Canada Setting: University hospital Providers: Unspecified OB/GYN Length of follow-up: None	No of subjects at start: 80 Drop-outs: 0 Loss to follow-up: NA No of subjects at end: 80 Inclusion criteria: Gestational age ≥ 36 weeks; Bishop score ≤ 4 Exclusion criteria: regular uterine contractions; contraindication to vaginal delivery; asthma or hypersensitivity to prostaglandins; prior attempts at ripening or induction in this pregnancy; malpresentation; multiple gestation; intrauterine death; polyhydramnios; antepartum hemorrhage; premature rupture of membranes; uterine scar Age (mean): 3 mg: 27.3 2 mg: 24.7 1 mg: 27.2 Placebo: 26.8 Race: NR Gestational age at entry (mean): 3 mg: 38.9 2 mg: 39.0 1 mg: 39.0 Placebo: 40.0 Dating criteria: NR Parity (% nulliparous): 3 mg: 40% 2 mg: 65% 1 mg: 65% Placebo: 55%	1) Change in Bishop score 2) Labor after gel alone 3) Cesarean section 4) Uterine hyper-contractility	1) Change in Bishop score: 3 mg: 3.8 2 mg: 2.6 1 mg: 2.7 Placebo: 1.4 $p < 0.01$ 2) Labor after gel alone: 3 mg: 50% 2 mg: 25% 1 mg: 5 % Placebo: 0% 3) Cesarean section: 3 mg: 20% 2 mg: 25% 1 mg: 35% Placebo: 15% 4) Uterine hypercontractility: 3 mg: 20% 2 mg: 10% 1 mg: 5% Placebo: 0%	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: - Statistical tests: - Gestational age: + Dating criteria: - Bishop score: + Underpowered to detect many important differences or trends. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
		Bishop score (mean): 3 mg: 2.6 2 mg: 3.0 1 mg: 2.7 Placebo: 2.4 Other: 18% of subjects induced for prolonged pregnancy			

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Grünberger and Spona, 1986	Design: RCT, method of randomization not described Interventions: 1) PGE_2 (1.5 mg) in saline (n = 15) Protocol: PGE_2 injected through syringe, using cervical cap. If labor within 6 hours, then cap removed; if no labor, then administration repeated. If no labor by 24 hours, then patient crossed over to other treatment group. Amniotomy performed when labor established and cervix sufficiently dilated (≥ 4 cm). 2) Placebo (n = 15) Protocol: Same as above, but with saline alone Dates: NR Location: Vienna, Austria Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 30 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 30 Inclusion criteria: 41-42 weeks gestation; unfavorable cervix Exclusion criteria: Maternal or fetal risk factors; twin pregnancy; breech presentation; previous C-section; previous surgery on cervix Age: NR Race: NR Gestational age at entry: NR (gestational age of 41-42 weeks required for entry into study) Dating criteria: NR Parity: Two groups “equal” (no further information provided) Bishop score (mean): PGE_2, 4.7; placebo, 4.6	1) Treatment failure (neither cervical ripening nor delivery)	1) Treatment failure: PGE_2: 1/15 (6.6%) Placebo: 10/15 (66.6%) $p < 0.001$	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: - Interventions described: + Mode of delivery: - Sample size: - Statistical tests: - Gestational age: - Dating criteria: - Bishop score: + Results summarized for period before crossover.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Gupta, Vasishta, Sawhney, et al., 1998	Design: RCT, randomization by computer-generated list and sealed envelope Interventions: 1) Stripping of membranes (n = 50) Protocol: Stripping of membranes performed at 38 weeks by digital separation of 2-3 cm of chorionic membranes from lower uterine segment using two circumferential passes of the examining fingers. Performed "under aseptic precautions." Patients then followed weekly (no details provided) until delivery or scheduled induction. 2) Gentle cervical exam (control) (n = 50) Protocol: Exam not described. Performed at 38 weeks "under aseptic precautions." Patients then followed weekly (no details provided) until delivery or scheduled induction. Dates: NR Location: Chandigarh, India Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 100 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 100 Inclusion criteria: Confirmed gestational age; early confirmation of pregnancy, cephalic presentation; no contraindication to vaginal delivery Exclusion criteria: Closed cervix at 38 weeks gestation; known medical disease or medical complications of pregnancy; multiple pregnancy; hydramnios; premature rupture of membranes; vaginal or cervical infection; low-lying placenta; intrauterine fetal death; malpresentation; labor; cephalopelvic disproportion Age (mean ± SD): Stripping, 24.46 ± 3.07; control, 23.52 ± 2.55 Race: NR Gestational age at entry (mean ± SD): Stripping, 38.00 ± 0.44 weeks; control, 38.02 ± 0.10 Dating criteria: NR Parity: 100% primigravidae Bishop score: Stripping, 86% < 6; control, 82% < 6	1) Apgar scores at 1 minute 2) Apgar scores at 5 minutes 3) Birthweight 4) Admission to NICU 5) Stillbirths 6) Gestational age at onset of labor 7) Days from intervention to delivery 8) Pregnancy continuing beyond 40 weeks 9) Induction of labor 10) C-sections 11) Assisted vaginal delivery 12) Microbiological flora	1) Apgar scores at 1 minute (mean ± SD): Stripping: 7.80 ± 0.17 Control: 7.74 ± 0.16 p > 0.05 2) Apgar scores at 5 minutes (mean ± SD): Stripping: 8.96 ± 0.19 Control: 9.12 ± 0.12 p > 0.05 3) Birthweight (mean ± SD): Stripping: 2882 ± 340 g Control: 2894 ± 420 g (no p-value reported) 4) Admission to NICU: Stripping: 0 Control: 2/50 (4%) (no p-value reported) 5) Stillbirths: Stripping: 1/50 (2%) Control: 0 p > 0.05 6) Gestational age at onset of labor (mean ± SD): Stripping: 38.70 ± 0.63 weeks Control: 39.83 ± 0.56 weeks p < 0.001 7) Days from intervention to delivery (mean ± SD): Stripping: 4.62 ± 4.15 Control: 11.95 ± 8.27 p < 0.005 8) Pregnancy continuing beyond 40 weeks: Stripping: 2/50 (4%) Control: 17/50 (34%)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				(no p-value reported) 9) Induction of labor: Stripping: 1/50 (2%) Control: 16/50 (32%) $p < 0.05$ 10) C-sections: *Overall:* Stripping: 6/50 (12%) Control: 8/50 (16%) $p > 0.05$ *For fetal distress:* Stripping: 3/50 (6%) Control: 5/50 (10%) (no p-value reported) *For nonprogress of labor:* Stripping: 3/50 (6%) Control: 3/50 (6%) (no p-value reported) 11) Assisted vaginal delivery: Stripping: 13/50 (26%) Control: 9/50 (18%) (no p-value reported) 12) Microbiological flora: No significant difference in the microbiological flora of cervical swabs (taken at time of intervention and at onset of labor) or the placental membrane in the two groups.	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Herabutya, Prasert-sawat, and Pokpirom, 1997	Design: RCT, blocked randomization scheme Interventions: 1) Misoprostol (n = 60) Protocol: 100-µg tablet placed in posterior vaginal fornix. 2) PGE_2 gel (n = 50) Protocol: PGE_2 gel (1.5 mg) placed via catheter into the endocervix In both groups, patients re-examined at 12 hours. Amniotomy carried out if cervix 80% effaced and 3 cm dilated. Patients who did not enter active labor or who had SROM without adequate uterine contractions were given oxytocin augmentation. At 24 hours, those still not in labor were sent to the labor ward for induction by amniotomy and oxytocin. Dates: May 1995 - Apr 1996 Location: Bangkok, Thailand Setting: University hospital Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: 110 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 110 Inclusion criteria: Medical or obstetric indication for induction; singleton pregnancy; cephalic presentation; intact membranes; Bishop score ≤ 4 Exclusion criteria: None specified Age (mean ± SD): Misoprostol, 29.12 ± 4.69; PGE_2, 28.18 ± 4.72 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 39.33 ± 1.41 weeks; PGE_2, 39.74 ± 1.43 weeks Dating criteria: NR Parity: Misoprostol, 73% nulliparous; PGE_2, 82% nulliparous Bishop score (mean ± SD): Misoprostol, 2.22 ± 1.06; PGE_2, 2.50 ± 1.15 Other: Indications for induction: Preeclampsia: 44% Postterm: 34% Decreased fetal movement: 9% Diabetes mellitus: 4% IUGR: 3% Previous dead fetus: 4% Nonreactive NST: 4%	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Time from induction to delivery 4) Hyperstimulation 5) C-sections	1) Apgar score < 7 at 1 minute: Misoprostol: 4/60 (6%) PGE_2: 4/50 (8%) p = 1.00 2) Apgar score < 7 at 5 minutes: Misoprostol: 0/60 PGE_2: 1/50 (2%) p = 0.45 3) Time from induction to delivery (mean ± SD): Misoprostol: 19.14 ± 10.64 hours PGE_2: 21.37 ± 13.09 hours p = 0.33 4) Hyperstimulation: Misoprostol: 1/60 PGE_2: 0/50 (no p-value reported) 5) C-sections: Misoprostol: 19/60 (31.7%) PGE_2: 16/50 (32.0%) p = 0.87	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (34% of total study population). Results not stratified by parity.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Howarth, Funk, Steytler, et al., 1996	Design: RCT, randomization by computer-generated list of random numbers and sealed envelopes Interventions: 1) Misoprostol (n = 36) Protocol: 100 µg misoprostol placed in posterior vaginal fornix. 2) PGE_2 gel (n = 36) Protocol: 1 mg PGE_2 gel placed in posterior vaginal fornix. In both groups, second dose administered after 6 hours if cervix remained unfavorable. Patients not in labor by 12 hours were managed according to their physician's preference. C-section was performed for suspected fetal distress. Dates: Apr - June 1995 Location: Pretoria, South Africa Setting: University hospital Type(s) of providers: General OB/GYN Length of follow-up: None	No. of subjects at start: 72 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 72 Inclusion criteria: Singleton pregnancy, longitudinal lie; cephalic presentation; fetal well-being; anticipated fetal mass > 2000 g; intact membranes; unfavorable cervix Exclusion criteria: Contra-indication to vaginal delivery; previous C-section; parity > 4; contraindication to prostaglandins Age (median, with range): Misoprostol, 27 (18-41); PGE_2, 27 (18-24) Race: NR Gestational age at entry (median, with range): Misoprostol, 40 weeks (35-43); PGE_2, 40 weeks (34-42) Dating criteria: NR Parity (median, with range): Misoprostol, 1 (0-4); PGE_2, 1 (0-4) Bishop score (median, with range): Misoprostol, 4 (2-7); PGE_2, 5 (2-7) Other: Indications for induction: Hypertension: 47% Postterm: 33% Other: 19%	1) Apgar score at 5 minutes 2) Birthweight 3) C-sections 4) Delivery within 12 hours 5) Tachysystole	1) Apgar score at 5 minutes (median, with range): Misoprostol: 10 (7-10) PGE_2: 10 (8-10) p = not significant 2) Birthweight (median, with range): Misoprostol: 3220 g (2260-4200) PGE_2: 2880 g (2100-4020) p = not significant 3) C-sections *Overall:* Misoprostol: 6/36 (17%) PGE_2: 15/36 (42%) $p < 0.05$ *For failed induction:* Misoprostol: 1/36 (3%) PGE_2: 6/36 (17%) p = not significant *For prolonged 1st stage of labor:* Misoprostol: 0/36 PGE_2: 7/36 (19%) $p < 0.01$ *For suspected fetal distress:* Misoprostol: 5/36 (14%) PGE_2: 2/36 (5.5%) p = not significant 4) Delivery within 12 hours: Misoprostol: 30/36 (83%) PGE_2: 13/36 (36%) $p < 0.05$ 5) Tachysystole: Misoprostol: 14/36 (39%) PGE_2: 3/36 (8%) $p < 0.01$	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (33% of total study population). Results not stratified by parity. 42% of patients in the misoprostol group were postdates vs. 25% in the PGE_2 group. Difference not significant, but study underpowered to detect differences at baseline or for outcomes.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Idrisa, Obisesan, and Adeleye, 1993	Design: RCT, patients assigned alternately to one of two treatment groups Interventions: 1) Membrane sweeping (n = 100) Protocol: Membrane sweeping performed at 41 weeks using the examiner's index finger. If no labor within 6 days, then patient induced with oxytocin. 2) Control (n = 100) Management of control group not specified Dates: Jan 1988 - Dec 1990 Location: Ibadan, Nigeria Setting: University hospital Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: 200 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 200 Inclusion criteria: Gestational age 41 weeks; no spontaneous labor Exclusion criteria: Contra-indications to vaginal delivery Age (mean ± SD): Membrane sweeping, 26 ± 3.1; control, 26 ± 3.3 Race: NR Gestational age at time of induction (mean ± SD): Both groups, 292 ± 2 days Dating criteria: 2nd trimester U/S Parity: NR Bishop score: NR	1) Birthweight 2) Perinatal death 3) Complications 4) Vacuum extraction/ forceps-assisted delivery 5) C-sections 6) Spontaneous labor	1) Birthweight (mean ± SD): Sweeping: 3.05 ± 0.25 kg Control: 3.05 ± 0.25 kg p = not significant 2) Perinatal death: Sweeping: 0/100 Control: 0/100 p = not significant 3) Complications: "No severe maternal or neonatal complication attributable to membrane sweeping was observed." 4) Vacuum extraction/ forceps-assisted delivery: Sweeping: 3/100 Control: 6/100 p = not significant 5) C-sections: Sweeping: 2/100 Control: 3/100 p = not significant 6) Spontaneous labor: Sweeping: 92/100 Control: 33/100 p < 0.001	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: - Results not stratified by parity.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Kadanali, Küçüköz-kan, Zor, et al., 1996	Design: RCT, randomization by sealed envelope Interventions: 1) Misoprostol (n = 112) Protocol: Misoprostol 100 µg tablet inserted intravaginally in the posterior fornix. Same dose repeated orally every 2 hours until adequate labor established (at least 3 contractions in 10 minutes). If labor not achieved by 24 hours, then patient infused with 10 IU oxytocin in 1000 ml 5% glucose solution. Infusion started at rate of 4 mIU/min and doubled every 30 minutes (to maximum of 32 mIU/min) until contractions began. If no active labor after 12 hours of oxytocin administration, then C-section performed. 2) PGE_2 gel + oxytocin (n = 112) Protocol: PGE_2 gel instilled into cervix. If no labor after 6 hours, then oxytocin infusion initiated "according to a uniform protocol." Dates: Mar-Aug 1995 Location: Erzurum, Turkey Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 224 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 224 Inclusion criteria: Medical or obstetrical indication for induction; no labor or fetal distress; gestational age 37-42 weeks; singleton vertex presentation Exclusion criteria: Previous uterine surgery, including C-section; Bishop score ≥ 6 Age (mean ± SD): Misoprostol, 22.3 ± 5.7; PGE_2/oxytocin, 22.5± 5.3 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 38.2 ± 3.4 weeks; PGE_2/oxytocin, 38.8 ± 2.8 weeks Dating criteria: NR Parity: Misoprostol, 70% nulliparous; PGE_2/oxytocin, 73% nulliparous Bishop score (mean ± SD): Misoprostol, 4.0 ± 1.4; PGE_2/oxytocin, 3.8 ± 1.4 Other: Indications for induction were as follows: Postdates: 41% Preeclampsia: 22% PROM: 11%	1) Apgar score < 5 at 5 minutes 2) Birthweight 3) Cord pH < 7.16 4) Vacuum extraction 5) C-sections for obstetric indication 6) C-sections for failed induction 7) Cost per patient 8) Time to delivery	1) Apgar score < 5 at 5 minutes: Misoprostol: 2/112 (1.8%) PGE_2/oxytocin: 2/112 (1.8%) p = not significant 2) Birthweight (mean ± SD): Misoprostol: 3382 ± 702.3 g PGE_2/oxytocin: 3302 ± 771.9 g p = not significant 3) Cord pH < 7.16: Misoprostol: 8/112 (7.1%) PGE_2/oxytocin: 10/112 (8.9%) p = not significant 4) Vacuum extraction: Misoprostol: 4/112 (3.6%) PGE_2/oxytocin: 5/112 (4.5%) p = not significant 5) C-sections for obstetric indication: Misoprostol: 5/112 (4.5%) PGE_2/oxytocin: 6/112 (5.4%) p = not significant 6) C-sections for failed induction: Misoprostol: 7/112 (6.3%) PGE_2/oxytocin: 15/112 (13.4%) p = 0.001 7) Cost per patient: Misoprostol: $1.50 PGE_2/oxytocin: $28.00 (no p-value reported) 8) Time to delivery (mean ± SD): Misoprostol: 9.2 ± 2.4 hours PGE_2/oxytocin: 15.2 ± 3.2 hours p = 0.001	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Mean gestational age 38 weeks, but 41% induced for "postdates." Results not reported separately for subgroup of patients induced for postterm pregnancy (41% of total study population). Results not stratified by parity. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
		Diabetes: 5% IUGR: 6% Other: 15%			
Kadar, Tapp, and Wong, 1990	Design: RCT, randomization by hospital number Interventions: 1) Nipple stimulation (NS) (n =62) Protocol: Women given written instructions for NS and instructed to perform unilateral NS manually each day "for as long as was practically feasible." Told to stop NS if contractions occurred more frequently than 5 in 10 minutes if a contractions lasted more than 90 seconds; NS could be resumed once the contractions had abated. 2) Control (no nipple stimulation) (n = 76) Dates: NR Location: London, England Setting: Outpatient clinic/physician office; university hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 155 Dropouts: 17 (11%) Loss to follow-up: NA No. of subjects at end: 138 Inclusion criteria: Low-risk pregnancy; ≥ 39 weeks gestation Exclusion criteria: None specified; patients withdrawn if pregnancy complications developed during the study Age (median): NS, 26.5; control, 25.0 Race: NS, 81% White; control, 75% White Gestational age at entry: Median, 281 days in both groups Dating criteria: LMP *or* U/S before 20 weeks Parity: NS, 52% nulliparous; control, 50% nulliparous Bishop score (median): Both groups, 5.0	1) Birthweight 2) Spontaneous delivery 3) Spontaneous labor 4) Postterm deliveries (> 294 days) 5) Pregnancy duration	1) Birthweight (median): NS: 3500 g Control: 3500 g (no p-value reported) 2) Spontaneous delivery: NS: 48/62 (77%) Control: 64/76 (84%) (no p-value reported) 3) Spontaneous labor: NS: 60/62 (97%) Control: 70/76 (92%) (no p-value reported) 4) Postterm deliveries: NS: 9/62 (14.5%) Control: 8/76 (10.5%) (no p-value reported) 5) Pregnancy duration (median): NS: 281 days Control: 281 days (no p-value reported)	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Compliance with nipple stimulation was poor. 70% of the women assigned to the NS group either failed to perform NS altogether or did so for < 2 hours in total. Survival analysis showed that duration of pregnancy was influenced only by the gestational age at enrollment and the Bishop score at enrollment. Nipple stimulation did not significantly affect the duration of pregnancy or the frequency of postterm deliveries. Women assigned to the nipple stimulation group who refused to participate were included with controls in the analysis.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Kemp, Winkler, and Rath, 2000	Design: RCT, randomization by stratified block Interventions: 1) PGE_2 vaginal gel (2 mg) (n = 229) Protocol: Gel administered in the posterior fornix. Repeated every 6-8 hours up to 3 times until Bishop score > 7. When Bishop score > 7, oxytocin administered 8 hours after last PGE_2 administration. If no labor and no improvement in Bishop score after 3 applications of gel, then 24-hour rest, followed by either induction with prostaglandins or C-section, as clinically indicated. FHR monitored for 2 hours following PGE_2 application and intermittently thereafter. 2) PGE_2 intracervical gel (0.5 mg) (n = 241) Protocol: Same as above, except that 0.5-mg gel administered "high into the cervical canal." Dates: Apr 1995 - July 1997 Location: Aachen, Germany Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 470 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 470 Inclusion criteria: Singleton pregnancy; vertex presentation; medical indication for induction (> 10 days postterm, premature rupture of the membranes, IUGR, hypertension; gestational or pre-existing diabetes); Bishop score 3-4 Exclusion criteria: Known contraindications for prostaglandins; previous uterine surgery; previous vertical C-section; uterine abnormality; FHR abnormality Age: NR Race: NR Gestational age at entry: NR; vaginal gel, 32.9% > 10 days postterm; intracervical gel, 29.2% > 10 days postterm Dating criteria: NR Parity: NR Bishop score: NR (required to be 3 or 4 for entry into study)	1) Apgar score ≤ 7 at 5 minutes 2) Umbilical artery pH < 7.20 3) C-sections 4) Change in Bishop score (before/after 1st administration) 5) Vaginal delivery within 24 hours 6) Time from induction to delivery 7) "Maternal side effects" 8) Hyperstimulation	1) Apgar score ≤ 7 at 5 minutes: Vaginal gel: 1.3% Intracervical gel: 2.1% p = not significant 2) Umbilical artery pH < 7.20: Vaginal gel: 12.3% Intracervical gel: 8.7% p = not significant 3) C-sections: Vaginal gel: 22.3% Intracervical gel: 26.7% p = not significant 4) Change in Bishop score (before/after 1st administration) (mean): Vaginal gel: 1.9 Intracervical gel: 1.35 p = 0.001 5) Vaginal delivery within 24 hours: Vaginal gel: 81.6% Intracervical gel: 67.8% p = 0.001 6) Time from induction to delivery (median): Vaginal gel: 15.7 hours Intracervical gel: 19.1 hours p = 0.01 7) "Maternal side effects": Vaginal gel: 5.7% Intracervical gel: 6.7% p = not significant 8) Hyperstimulation: Vaginal gel: 14.5% Intracervical gel: 13.0% p = not significant	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: - Dating criteria: - Bishop score: +

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Kramer, Gilson, Morrison, et al., 1997	Design: RCT, randomization schedule computer-generated by hospital pharmacy Interventions: 1) Misoprostol (100 µg) (n = 60) Protocol: Misoprostol 100 µg placed in posterior vaginal fornix every 4 hours until adequate uterine contractions achieved (defined as > 200 Montevideo units). No lubricating gel used to place tablets. Repeat dosing (up to max of 5 doses) permitted if uterine activity inadequate and fetus tolerating labor. Oxytocin started if labor had not progressed by 4 hours after last dose of misoprostol. 2) Oxytocin infusion (n = 66) Protocol: Intravenous oxytocin started at an infusion rate of 1 mU/min. Dose increased every 30 min until adequate uterine activity achieved (> 200 Montevideo units). Maximal infusion rate permitted was 36 mU/min. Women in both groups were monitored by external tocodynamometry. Fetal scalp monitoring, cord blood gas sampling, and administration of terbutaline left to discretion of managing physician. Amniotomy generally performed at 3-4 cm dilation. Dates: June 1995 - Apr 1996	No. of subjects at start: 130 Dropouts: 4 women excluded from analysis after randomization Loss to follow-up: NA No. of subjects at end: 126 Inclusion criteria: None stated Exclusion criteria: Multiple gestation; nonvertex presentation; abnormal FHR tracing; previous uterine surgery; allergy to misoprostol; history of asthma; digital exam with lubricant immediately before induction; spontaneous uterine contractions more frequently than every 5 minutes; contraindications to vaginal delivery (e.g., active genital herpes, placenta previa) Age (mean ± SD): Misoprostol, 26.2 ± 5.9; oxytocin, 25.4 ± 5.7 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 39.6 ± 2.6 weeks; oxytocin, 38.3 ± 3.2 weeks Dating criteria: NR Parity: Misopr / Oxytocin Nulliparous 60% / 49% Primiparous 20% / 29% Multiparous 20% / 22% Bishop score (% with score ≤ 3): Misoprostol, 58%; oxytocin, 38% Other: Indications for induction	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Arterial cord blood pH 4) Birthweight 5) Vacuum delivery 6) Forceps delivery 7) C-section for nonreassuring FHR tracing 8) C-section for dystocia 9) C-section for worsening maternal status 10) Duration of labor 11) Tachystole 12) Estimated hospital charges	1) Apgar score < 7 at 1 minute: Misoprostol: 8/60 (13%) Oxytocin: 12/66 (18%) p = not significant 2) Apgar score < 7 at 5 minutes: Misoprostol: 0/60 Oxytocin: 3/66 (5%) p = not significant 3) Arterial cord blood pH (mean ± SD): Misoprostol (n = 16): 7.21 ± 0.08 Oxytocin (n = 9): 7.19 ± 0.16 p = not significant 4) Birthweight (mean ± SD): Misoprostol: 3262 ± 679 g Oxytocin: 3092 ± 786 p = not significant 5) Vacuum delivery: Misoprostol: 2/60 (3%) Oxytocin: 3/66 (5%) p = not significant 6) Forceps delivery: Misoprostol: 6/60 (10%) Oxytocin: 6/66 (9%) p = not significant 7) C-section for nonreassuring FHR tracing: Misoprostol: 7/60 (12%) Oxytocin: 4/66 (6%) p = not significant 8) C-section for dystocia: Misoprostol: 6/60 (10%) Oxytocin: 14/66 (21%) p < 0.05 9) C-section for worsening maternal status: Misoprostol: 0/60	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (29% of total study population). Results not stratified by parity. 4/60 women in the misoprostol group received oxytocin, but were analyzed in intention-to-treat fashion as part of the misoprostol group. Difference in baseline characteristics suggests problem with randomization. Underpowered to detect some differences in baseline and other variables. Sample size estimates based on time to delivery. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Location: Albuquerque, New Mexico Setting: University hospital Type(s) of providers: General OB/GYN resident physicians under direct supervision of faculty member Length of follow-up: None	were as follows: Preeclampsia: 41% Postterm: 29% Oligohydramnios: 11% Diabetes mellitus: 2% Fetal growth restriction: 1% Other: 16%		Oxytocin: 1/66 (2%) p = not significant 10) Duration of labor (median, with range): Misoprostol: 585 minutes (120-1890) Oxytocin: 885 minutes (120-1890) p < 0.001 11) Tachystole: Misoprostol: 42/60 (70%) Oxytocin: 7/66 (11%) p < 0.001 12) Estimated hospital charges (total charges per patient [mean ± SD]): Misoprostol: $2081 ± $984 Oxytocin: $2616 ± $1035 p < 0.005	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Lee, 1997	Design: RCT, randomization by sealed envelope Interventions: 1) Misoprostol (n = 25) Protocol: 200 µg given intravaginally at 6-hour interval up to a maximum of 2 doses. Patient examined every 6 hours and transferred to labor room when "ready for labor." If no established labor, then oxytocin given. If cervix still unripe after 24 hours, then C-section performed. 2) PGE_2 (n = 25) Protocol: Same as above, except that PGE_2 3 mg used instead of misoprostol. Dates: Beginning Jan 1996 (no end date specified) Location: Pahang, Malaysia Setting: Community hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 50 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 50 Inclusion criteria: At least term + 10 days' gestation; para ≤ 3; singleton pregnancy; cephalic presentation; no prior C-section; no contraindication to prostaglandins; uncomplicated gestation; Bishop score ≤ 6 Exclusion criteria: None specified Age (mean ± SD): Misoprostol, 26.3 ± 4.8; PGE_2, 26.5 ± 4.4 Race: Misoprostol, 84% Malay; PGE_2, 72% Malay Gestational age at entry (mean number of days postdate [± SD]): Misoprostol, 12.5 ± 2.1 days; PGE_2, 12.6 ± 2.6 days Dating criteria: NR Parity (mean ± SD): Misoprostol, 1.3 ± 1.2; PGE_2, 1.1 ± 1.0 Bishop score (mean ± SD): Misoprostol, 4.1 ± 1.1; PGE_2, 4.1 ± 1.2	1) Apgar score at 1 minute 2) Apgar score at 5 minutes 3) Neonatal complication 4) Neonatal hospital stay 5) Moderate meconium aspiration 6) Established labor rate 7) Time to delivery 8) C-sections 9) Polysystole	1) Apgar score at 1 minute (mean ± SD): Misoprostol: 7.7 ± 0.7 PGE_2: 7.6 ± 1.3 p = 0.69 2) Apgar score at 5 minutes (mean ± SD) Misoprostol: 8.9 ± 0.4 PGE_2: 8.7 ± 1.1 p = 0.39 3) Neonatal complication: Misoprostol: 4/25 (16%) PGE_2: 1/25 (4%) p = 0.17 4) Neonatal hospital stay (mean ± SD): Misoprostol: 2.9 ± 2.3 days PGE_2: 2.7 ± 1.0 days p = 0.69 5) Moderate meconium aspiration: Misoprostol: 2/25 (8%) PGE_2: 1/25 (4%) (no p-value reported) 6) Established labor rate: Misoprostol: 23/25 (92%) PGE_2: 16/25 (64%) p = 0.04 7) Time to delivery: *Mean ± SD:* Misoprostol: 676.1 ± 411 minutes PGE_2: 874.9 ± 406 minutes p = 0.09 *Delivered by 6 hours:* Misoprostol: 5/25 (20%) PGE_2: 3/25 (12%) p = 0.35	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not stratified by parity. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				Delivered by 12 hours: Misoprostol: 18/25 (72%) PGE_2: 7/25 (28%) p = 0.047 8) C-sections: Misoprostol: 2/25 (8%) PGE_2: 4/25 (16%) p = 0.33 9) Polysystole: Misoprostol: 7/25 (28%) PGE_2: 3/25 (12%) p = 0.28	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Lien, Morgan, Garite, et al., 1998	Design: RCT, randomization by computer-generated table of random numbers Interventions: 1) PGE_2 gel (n = 43) Protocol: PGE_2 gel administered into the endocervical canal. Patient monitored continuously for ≥ 40 minutes. If FHR monitoring "reassuring," then patient instructed to return in 3-4 days for another NST, AFI determination, and gel insertion (up to maximum of 4 doses). Patient induced at 42 weeks, or before then if Bishop score > 9 or "an obstetric factor other than postdate pregnancy developed." Obstetric management during labor determined by patient's obstetrician. 2) Placebo gel (n = 47) Protocol: Same as above, except that identical placebo gel used instead of PGE_2 gel. Dates: NR Location: Anaheim, CA, and Portland, OR Setting: 1 university hospital and 3 community hospitals Type(s) of providers: General OB/GYN; nurse midwives Length of follow-up: None	No. of subjects at start: 92 Dropouts: 2 Loss to follow-up: NA No. of subjects at end: 90 Inclusion criteria: Gestational age ≥ 40 weeks, 3 days; Bishop score ≤ 6; AFI > 5 cm; reactive NST Exclusion criteria: Evidence of hyperstimulation; suspicious FHR patterns; ≥ 5 previous deliveries; nonvertex presentation; multiple gestation; previous C-section; major uterine surgery; placenta previa; other contraindications to vaginal delivery Age (mean ± SD): PGE_2, 25.9 ± 7.0; placebo, 26.4 ± 5.8 Race: PGE_2: 84% White, 12% Hispanic, 5% Asian/Black/other; placebo: 85% White, 6% Hispanic, 9% Asian/Black/other Gestational age at entry (mean ± SD): PGE_2, 40.9 ± 0.3 weeks; placebo, 40.7 ± 0.3 weeks (p = 0.01) Dating criteria: LMP confirmed by either 1st trimester pelvic exam or U/S before 24 weeks Parity: PGE_2, 67% nulliparous; placebo, 55% nulliparous Bishop score (median, with range): PGE_2, 3 (1-6); placebo, 3 (0-5)	1) Apgar score ≤ 7 at 5 minutes 2) Birthweight (mean) 3) Birthweight > 4000 g 4) Shoulder dystocia 5) Gestational age at delivery 6) Time from enrollment to delivery 7) C-sections 8) Vacuum- or forceps-assisted delivery 9) Chorioamnionitis 10) Endometritis	1) Apgar score ≤ 7 at 5 minutes: PGE_2: 0 Placebo: 1/47 (2.1%) p = not significant 2) Birthweight (mean ± SD): PGE_2: 3765 ± 446 Placebo: 3684 ± 411 p = not significant 3) Birthweight > 4000 g: PGE_2: 14/43 (32.6%) Placebo: 7/47 (14.9%) p < 0.05 4) Shoulder dystocia: PGE_2: 3/43 (7.0%) Placebo: 1/47 (2.1%) p = not significant 5) Gestational age at delivery (mean ± SD): PGE_2: 41.7 ± 0.5 weeks Placebo: 41.6 ± 0.4 weeks p = not significant 6) Time from enrollment to delivery (mean ± SD): PGE_2: 5.5 ± 3.5 days Placebo: 6.0 ± 2.8 days p = not significant 7) C-sections: Overall: PGE_2: 6/43 (14.0%) Placebo: 8/47 (17.0%) p = not significant For fetal distress: PGE_2: 0 Placebo: 1/47 (2.1%) p = not significant *(continued on next page)*	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Results not stratified by parity.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				8) Vacuum- or forceps-assisted delivery: PGE_2: 6/43 (14.0%) Placebo: 3/47 (6.4%) p = not significant 9) Chorioamnionitis: PGE_2: 5/43 (11.6%) Placebo: 2/47 (4.3%) p = not significant 10) Endometritis: PGE_2: 1/43 (2.3%) Placebo: 1/47 (2.1%) p = not significant	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
MacKenzie and Burns, 1997	Design: RCT, randomization by computer-generated random numbers and sealed envelopes Interventions: 1) PGE_2 gel, 1 dose (n = 483) Protocol: PGE_2 gel (2 mg) applied vaginally. If labor had not started 14-20 hours after initial treatment, then amniotomy performed and IV oxytocin infusion started 1-2 hours later. If amniotomy not technically possible, it was deferred until 4 hours after oxytocin started. 2) PGE_2 gel, 2 doses (n = 472) Protocol: Same as above, but second dose of PGE_2 gel applied 6 hours after the first if labor not established or cervical score < 9. Dates: NR Location: Oxford, England Setting: Unspecified hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 1000 Dropouts: 45 (excluded due to protocol violations) Loss to follow-up: NA No. of subjects at end: 955 Inclusion criteria: Modified Bishop score ≤ 8; singleton viable pregnancy; cephalic presentation; no previous C-section Exclusion criteria: None specified Age: 1 dose / 2 doses < 20: 5% / 5% 20-29: 57% / 60% 30-39: 36% / 34% ≥ 40: 2% / 1% Race: NR Gestational age at entry (weeks): 1 dose / 2 doses < 40: 21% / 22% 40-42: 74% / 72% > 42: 5% / 6% Dating criteria: NR Parity: 1 dose / 2 doses 0: 49% / 55% 1-2: 46% / 39% ≥ 3: 5% / 6% Bishop score: 1 dose / 2 doses < 4: 25% / 29% 4-5: 44% / 39% ≥ 6: 31% / 31%	1) Apgar score < 8 at 1 minute 2) Apgar score < 5 at 1 minute 3) Apgar score < 8 at 5 minutes 4) Apgar score < 9 at 10 minutes 5) Birthweight 6) Admission to NICU 7) C-sections 8) Time to delivery	1) Apgar score < 8 at 1 minute: 1-dose nulliparae: 38/237 (16%) 2-dose nulliparae: 63/262 (24%) 1-dose multiparae: 43/246(17%) 2-dose multiparae: 37/210 (18%) (no p-value reported) 2) Apgar score < 5 at 1 minute: 1-dose nulliparae: 9/237 (4%) 2-dose nulliparae: 15/262 (6%) 1-dose multiparae: 15/246 (6%) 2-dose multiparae: 7/210 (3%) (no p-value reported) 3) Apgar score < 8 at 5 minutes: 1-dose nulliparae: 1/237 (< 1%) 2-dose nulliparae: 7/262 (3%) 1-dose multiparae: 5/246 (2%) 2-dose multiparae: 3/210 (1%) (no p-value reported) 4) Apgar score < 9 at 10 minutes: 1-dose nulliparae: 0/237 2-dose nulliparae: 3/262 (1.2%) 1-dose multiparae: 2/246 (0.8%) 2-dose multiparae: 0/210 (no p-value reported) 5) Birthweight (mean ± SD): 1-dose nulliparae: 3499 ± 546 g 2-dose nulliparae: 3512 ± 508 g p = 0.783 1-dose multiparae: 3646 ± 483 g 2-dose multiparae: 3642 ± 542 g p = 0.934 6) Admission to NICU: 1-dose nulliparae: 4/237 (2%) 2-dose nulliparae: 13/262 (5%) 1-dose multiparae: 6/246 (2%) 2-dose multiparae: 6/210 (3%) (no p-value reported)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (68% of total study population).

(continued on next page)

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
		Other: Indications for induction: Postterm: 68% Hypertension: 15% Fetal concerns: 6% Maternal health concerns: 1% Maternal request: 8% Past obstetric history: 2%		7) C-sections: 1-dose nulliparae: 35/237 (15%) 2-dose nulliparae: 30/262 (11%) RR = 1.0 (95% CI, 0.90-1.03) 1-dose multiparae: 4/246 (2%) 2-dose multiparae: 5/210 (2%) RR = 0.7 (95% CI, 0.19-2.51) 8) Time to delivery (mean ± SD): 1-dose nulliparae: 1240 ± 540 minutes 2-dose nulliparae: 1197 ± 503 minutes p = 0.358 1-dose multiparae: 927 ± 519 minutes 2-dose multiparae: 785 ± 394 minutes p = 0.001	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Magann, Chauhan, Nevils, et al., 1998	Design: RCT, randomization by table of random numbers and sealed envelopes Interventions: 1) PGE_2 gel (n = 35) Protocol: PGE_2 gel (0.5 mg) placed into cervix on a daily basis. Modified biophysical profile performed, and patient sent home only when monitoring revealed that any contractions caused had begun to dissipate. Labor induced when Bishop score = 8 or when patient reached 42nd week of pregnancy. 2) Membrane stripping (n = 35) Protocol: Membrane stripping performed daily + modified biophysical profile every 3 days. Membranes separated from the lower uterine segment by two circumferential sweeps of examining finger. If cervix unfavorable for stripping, it was stretched by examining finger daily until membrane stripping could be accomplished. Labor induced when Bishop score = 8 or when patient reached 42nd week of pregnancy. 3) Cervical exam (control) (n = 35) Protocol: Gentle cervical exam performed daily + modified biophysical profile every 3 days. Labor induced when Bishop score = 8 or when patient reached 42nd	No. of subjects at start: 105 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 105 Inclusion criteria: ≥ 41 weeks gestation; uncomplicated pregnancy; no contraindications to vaginal delivery; Bishop score ≤ 4 Exclusion criteria: None specified Age (mean ± SD): PGE_2: 24.5 ± 5.2 Stripping: 25.1 ± 5.1 Control: 25.5 ± 5 Race: PGE_2: 71% White, 11% Black, 17% Hispanic Stripping: 74% White, 11% Black, 11% Hispanic, 3% Asian Control: 63% White, 20% Black, 11% Hispanic, 6% Asian Gestational age at entry (mean ± SD): All 3 groups, 41.1 ± 0.1 weeks Dating criteria: LMP, early pelvic exam, auscultation of the fetal heart by U/S stethoscope, and ("in nearly all cases") U/S before 20th week Parity: PGE_2: 0, 74%; 1, 14%; ≥ 2, 11% Stripping: 0, 51%; 1, 31%; ≥ 2, 17% Control: 0, 60%; 1, 26%; ≥ 2, 14%	1) Apgar score < 7 at 5 minutes 2) Birthweight 3) Uterine artery pH (mean) 4) Uterine artery pH < 7.2 5) Admitted to well-baby nursery 6) Inductions at 42 weeks 7) C-sections 8) Forceps-assisted deliveries 9) Total antepartum costs (per group) 10) Total intrapartum costs (per group)	1) Apgar score < 7 at 5 minutes: PGE_2: 1/35 (3%) Stripping: 0 Control: 1/35 (3%) p = 0.6 2) Birthweight (mean ± SD): PGE_2: 3694 ± 419 g Stripping: 3835 ± 489 g Control: 3770 ± 430 g p = 0.19 3) Uterine artery pH (mean ± SD): PGE_2: 7.22 ± 0.05 Stripping: 7.22 ± 0.05 Control: 7.21 ± 0.05 p = 0.77 4) Uterine artery pH < 7.2: PGE_2: 7/35 (20%) Stripping: 8/35 (23%) Control: 7/35 (20%) p = 0.94 5) Admitted to well-baby nursery: PGE_2: 32/35 (91%) Stripping: 33/35 (94%) Control: 35/35 (100%) p = 0.23 6) Inductions at 42 weeks: PGE_2: 7/35 (20%) Stripping: 6/35 (17%) Control: 22 (63%) p < 0.0001 7) C-sections: PGE_2: 8/35 (23%) Stripping: 5/35 (14%) Control: 5/35 (14%) (no p-value reported)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Sample size estimates based on post hoc analysis of proportion of patients induced at 42 weeks. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	week of pregnancy. Dates: Mar-Sep 1996 Location: San Diego, CA Setting: Military hospital Type(s) of providers: General OB/GYN Length of follow-up: None	Bishop score (mean ± SD): PGE_2: 2.6 ± 1 Stripping: 2.8 ± 0.7 Control: 2.6 ± 0.7		8) Forceps-assisted deliveries: PGE_2: 3/35 (9%) Stripping: 4/35 (11%) Control: 5/35 (14%) (no p-value reported) 9) Total antepartum charges (per group): PGE_2: $30,800 Stripping: $7420 Control: $9520 (no p-value reported) 10) Total intrapartum charges (per group): PGE_2: $11,445 Stripping: $9240 Control: $14,735 (no p-value reported)	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Magann, McNamara, Whitworth, et al., 1998	Design: RCT, randomization by table of random numbers and sealed envelopes Interventions: 1) Membrane stripping (n = 33) Protocol: Membrane stripping performed every 3 days by placing a finger through the cervix and performing 2 circumferential sweeps. If the cervix would not admit a finger, then examining finger placed into the cervix every 3 days until the sweeping could be performed. 3) Vaginal exam (control) (n = 32) Protocol: Gentle vaginal exam performed every 3 days. In both groups, treatment continued until spontaneous labor, rupture of the membranes, or completion of 41 weeks' gestation, at which time patient admitted for induction of labor. Dates: NR Location: San Diego, CA, and Jackson, MS Setting: 1 university hospital and 1 military hospital Type(s) of providers: Not specified Length of follow-up: None	No. of subjects at start: 65 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 65 Inclusion criteria: 39 weeks' gestation; negative fetal fibronectin test result; Bishop score ≤ 4; vertex presentation Exclusion criteria: Placenta previa; other contraindications to vaginal delivery Age (mean ± SD): Stripping, 24.5 ± 5; control, 24.3 ± 5.3 Race: Stripping: 64% White, 27% Black, 9% Hispanic Control: 66% White, 22% Black, 6% Hispanic, 6% other Gestational age at entry (mean ± SD): Both groups, 39.00 ± 0.00 Dating criteria: LMP, initial exam, first auscultation of fetal heart tones with an U/S stethoscope, or U/S before 20 weeks Parity: Stripping, 55% nulliparous; control, 56% nulliparous Bishop score (mean ± SD): Stripping, 2.5 ± 0.6; control, 2.6 ± 0.9	1) Birthweight 2) Umbilical artery pH 3) Admission to NICU 4) Gestational age at delivery 5) Bishop score ≥ 8 at delivery 6) Inductions at 42 weeks 7) C-sections 8) Time from admission to delivery	1) Birthweight (mean ± SD): Stripping: 3449 ± 442 g Control: 3531 ± 490 g p = 0.48 2) Umbilical artery pH (mean ± SD): Stripping: 7.24 ± 0.04 Control: 7.23 ± 0.06 p = 0.43 3) Admission to NICU: Stripping: 2/33 (6%) Control: 2/32 (6%) p =1.00 4) Gestational age at delivery (mean ± SD): Stripping: 39.9 ± 0.3 weeks Control: 41.5 ± 0.6 weeks p < 0.0001 5) Bishop score ≥ 8 at delivery: Stripping: 19/33 (58%) Control: 6/32 (19%) p = 0.0002 6) Inductions at 42 weeks: Stripping: 0 Control: 18/32 (56%) p < 0.0001 7) C-sections: Stripping: 4/33 (12%) Control: 5/33 (15%) p = not significant 8) Time from admission to delivery (mean ± SD): Stripping: 10.4 ± 5.5 hours Control: 13.0 ± 7.1 hours p = 0.10	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Sample size estimates based on reduction in 42-week inductions.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Mahmood, 1989	Design: RCT, randomization by sealed envelope Interventions: 1) PGE_2 gel 2 mg (n = 40), inserted into posterior fornix at 5 PM day before induction; patients monitored for 1 hour after insertion. Second dose if Bishop score < 5 next morning at 9 AM. If no labor or cervical change by 9 AM next day, third insertion. 2) PGE_2 3 mg tablet (n = 40), inserted into posterior fornix. Protocol same as above. Dates: NR Location: Abderdeen, UK Setting: Community hospital Type(s) of providers: NR Length of follow-up: None	No of subjects at start: 80 Drop-outs: 0 Loss to follow-up: NA No of subjects at end: 80 Inclusion criteria: Gestational age 37-43 weeks; singleton pregnancy; vertex presentation; Bishop score < 5 Exclusion criteria: None specified Age (mean ± SD): Gel: 25 ± 4.4 Tablet: 25 ± 5.3 Race: NR Gestational age at entry: NR Dating criteria: NR Parity: NR Bishop score (mean and range): Gel: 2.30 (0-4) Tablet: 2.55 (0-4) Other: 61% induced for prolonged pregnancy	1) Number of insertions required for spontaneous labor 2) Time from insertion to spontaneous labor 3) Posttreatment Bishop score 4) Need for oxytocin 5) Emergent cesarean section 6) Apgar score < 7 at 1 minute 7) Apgar score < 7 at 5 minutes	1) Number of insertions required for spontaneous labor: Gel: 1 insertion: 50% 2 insertions: 50% Tablet: 1 insertion: 20% 2 insertions: 50% 3 insertions: 30% p < 0.05 2) Time from insertion to onset of labor: Gel: 15.1, if spontaneous; 20.6, if induction needed Table: 25.6, if spontaneous; 30.5, if induction needed p < 0.02 3) Posttreatment Bishop score: Gel: 9.5 Tablet: 7.0 p < 0.05 4) Need for oxytocin: Gel: 12.5% Tablet: 50% p < 0.001 5) Emergent cesarean section: Gel: 15% Tablet: 30% p = not significant 6) Apgar score < 7 at 1 minute: Gel: 22% Tablet: 37% 7) Apgar score < 7 at 5 minutes: Gel: 0 Tablet: 2.5%	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: - Statistical tests: - Gestational age: - Dating criteria: - Bishop score: +

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
McColgin, Hampton, McCaul, et al., 1990	Design: RCT, randomization by computer Interventions: 1) Membrane stripping (n = 90) Protocol: Performed weekly by digital separation of 2-3 cm of the membranes from the lower uterine segment using 2 circumferential passes of the examining finger. If cervix long and closed, then stretched digitally until membrane stripping could be accomplished. Treatment continued until patient admitted to labor and delivery or advanced beyond 42 completed weeks' gestation. 2) Cervicovaginal exam (control) (n = 90) Protocol: Weekly atraumatic assessment of the cervix for Bishop scoring. Treatment continued until patient admitted to labor and delivery or advanced beyond 42 completed weeks' gestation. Dates: Mar 1988 - June 1989 Location: Jackson, MS Setting: University hospital Type(s) of providers: General OB/GYN Length of follow-up: None	No. of subjects at start: 209 Dropouts: 29 (excluded post-randomization Loss to follow-up: NA No. of subjects at end: 180 Inclusion criteria: 38 weeks' gestation Exclusion criteria: Placenta previa; low-lying placenta; abnormal fetal presentation; known medical complication; vaginal or cervical infection Age (mean ± SEM): Stripping, 23.06 ± 0.55; control, 23.31 ± 0.58 Race: NR Gestational age at entry: 38 weeks Dating criteria: LMP, early assessment of uterine size, and U/S before 20 weeks Parity: Stripping, 40% nulliparous; control, 50% nulliparous Bishop score (mean ± SEM): Stripping, 3.51 ± 0.24; control, 3.82 ± 0.19 Other: Long/closed cervix: Stripping, 12/90 (13%); control, 10/90 (11%)	1) Fetal deaths 2) Delivery ≥ 42 weeks 3) Days to delivery 4) Delivery within 1 week	1) Fetal deaths: Stripping: 0 Control: 1/90 (1%) (no p-value reported) 2) Delivery ≥ 42 weeks: Stripping: 3/90 (3.3%) Control: 14/90 (15.6%) p < 0.004 3) Days to delivery (mean ± SEM): Stripping: 8.60 ± 0.74 Control: 15.14 ± 0.83 p < 0.001 4) Delivery within 1 week: Stripping: 49/90 (54.5%) Control: 14/90 (15.6%) p < 0.001	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: NA Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Investigators stated that "nulliparous patients and individuals with unfavorable Bishop scores benefited the most from membrane stripping in reduction of postterm pregnancies." No quantitative data provided.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
McColgin, Patrissi, and Morrison, 1990	Design: RCT, method of randomization not specified Interventions: 1) Sweeping membranes (n = 51) performed weekly from 38-42 weeks by digital separation of membranes from lower uterine segment; if cervix closed, "digitally stretched" to allow sweeping. 2) Control (n = 48): Bishop scoring only performed weekly from 38-42 weeks Both groups followed until 42 weeks, when induction scheduled Dates: NR Location: Jackson, MS Setting: Military hospital and university hospital antenatal clinics Type(s) of providers: NR Length of follow-up: None	No of subjects at start: 103 Drop-outs: 4 Loss to follow-up: NA No of subjects at end: 99 Inclusion criteria: Low-risk pregnancy Exclusion criteria: Uncertain dates; abnormal presentation; low-lying placenta; scheduled repeat cesarean; candidates for vaginal birth after cesarean section allowed to participate Age: NR ("comparable") Gestational age at entry: NR ("comparable") Dating criteria: LMP and ultrasound prior to 20 weeks Parity: NR ("comparable") Bishop score: NR ("comparable")	1) Days to delivery 2) Proportion delivering within 1 week 3) Number delivering after 42 weeks 4) Cesarean delivery	1) Days to delivery: Sweeping: 6.7 days Control: 13.3 days p = 0.003 2) Proportion delivering within 1 week: Sweeping: 59% Control: 21% p = 0.003 3) Number delivering after 42 weeks: Sweeping: 2 Control: 6 p = 0.12 4) Cesarean delivery: Sweeping: 7/51 Control: 5/48 p = NS Results similar when analysis restricted to those entering study at 38 weeks. No significant differences seen in group with Bishop score > 5.	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: - Dating criteria: + Bishop score: -

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Misra and Vavre, 1994	Design: RCT, method of randomization not described Interventions: 1) Intracervical PGE_2 gel (0.5 mg) (n = 136) Protocol: Gel administered into cervical canal at 7:30 PM the night before induction. If no labor (3-4 "good intensity" contractions, lasting 40-50 seconds each, every 10 minutes) after 12 hours, then patient induced with oxytocin. Amniotomy performed after cervical dilatation of ≥ 2.5 cm and effacement of ≥ 80%. If no labor after 12 hours and after receiving as much as 64 mU/min of oxytocin, then C-section performed. 2) Oxytocin infusion (n = 127) Protocol: Infusion started at 8:00 AM on day of planned induction, beginning with 2 mU/min and increasing the dose by 1-2 mU every 30 minutes. Amniotomy performed after cervical dilatation of ≥ 2.5 cm and effacement of ≥ 80%. If no labor after 12 hours and after receiving as much as 64 mU/min of oxytocin, then C-section performed. Dates: Aug 1992 - Jan 1994 Location: Bhilai, India Setting: Unspecified hospital Type(s) of providers:	No. of subjects at start: 263 Dropouts: 0 Loss to follow-up: 0 No. of subjects at end: 263 Inclusion criteria: Bishop score < 4; induction of labor required Exclusion criteria: Premature rupture of membranes; "major degrees of cephalopelvic disproportion"; malpresentations; intrauterine deaths; congenital anomalies not compatible with life; persistently nonreactive NST Age (mean ± SD): PGE_2 primigravidas (n =80): 23.7 ± 3.7 PGE_2 multigravidas (n = 56): 25.6 ± 4.0 Oxytocin primigravidas (n = 72): 23.3 ± 2.4 Oxytocin multigravidas (n = 55): 26.3 ± 3.3 Race: NR Gestational age at entry: PGE_2 primigravidas: 39.6 ± 2.7 weeks PGE_2 multigravidas: 39.4 ± 2.1 weeks Oxytocin primigravidas: 39.8 ± 2.0 weeks Oxytocin multigravidas: 39.5 ± 2.3 weeks Dating criteria: NR Parity: PGE_2, 59% primigravidas;	1) Apgar score < 7 at 5 minutes 2) Birthweight 3) Forceps/ventouse deliveries 4) C-sections	1) Apgar score < 7 at 5 minutes: PGE_2 primigravidas: 3/80 (3.8%) Oxytocin primigravidas: 2/72 (2.8%) p = not significant PGE_2 multigravidas: 1/56 (1.8%) Oxytocin multigravidas: 0 p = not significant 2) Birthweight (mean ± SD): PGE_2 primigravidas (n = 80): 2640 ± 580 g Oxytocin primigravidas (n = 72): 2660 ± 550 g p = 0.84 PGE_2 multigravidas (n = 56): 2670 ± 580 g Oxytocin multigravidas (n = 55): 2770 ± 620 g p = 0.38 3) Forceps/ventouse deliveries: PGE_2 primigravidas: 3/80 (3.8%) Oxytocin primigravidas: 4/72 (5.6%) (no p-value reported) PGE_2 multigravidas: 0 Oxytocin multigravidas: 2/55 (3.6%) (no p-value reported) 4) C-sections: PGE_2 primigravidas: 21/80 (26.3%) Oxytocin primigravidas: 34/72 (47.2%) p < 0.01 PGE_2 multigravidas: 7/56 (12.5%) Oxytocin multigravidas: 8/55 (14.6%) p = 0.75	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: - Interventions described: + Mode of delivery: - Sample size: - Statistical tests: - Gestational age: + Dating criteria: - Bishop score: + *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Unspecified OB/GYN Length of follow-up: None	oxytocin, 57% primigravidas Bishop score (mean ± SD): PGE_2 primigravidas: 2.2 ± 0.6 PGE_2 multigravidas: 2.5 ± 0.6 Oxytocin primigravidas: 2.3 ± 0.6 Oxytocin multigravidas: 2.6 ± 0.7			

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Mundle and Young, 1996	Design: RCT, randomization by random-number tables and sealed envelopes Interventions: 1) Misoprostol (n = 111) Protocol: 50-µg tablet placed in upper vagina every 4 hours until patient experienced progressive labor, contractions 3 times/minute, ruptured membranes, non-reassuring FHR tracing, or delivery. No more than 16 applications permitted; no change in dosage permitted. 2) PGE_2 gel (n = 111) Protocol: Patient given PGE_2 gel in dose of either 0.5 mg intracervically (for ripening) or 1-2 mg intravaginally (for induction), as determined by treating physician. In both groups, amniotomy was performed at the discretion of the attending physician. Oxytocin administration was begun at 2 mU/min, then increased by 2-mU/min increments at 30-60-min intervals. Oxytocin not permitted within 4 hours of last dose of misoprostol or 6 hours of last dose of PGE_2 gel. Dates: Mar-Sep 1994 Location: St. John's, Newfoundland, Canada Setting: Unspecified hospital	No. of subjects at start: 222 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 222 Inclusion criteria: Indication for induction; single live fetus; gestational age > 37 weeks; cephalic presentation; intact membranes Exclusion criteria: Nonreassuring FHR tracing; prior uterine surgery; know hypersensitivity to misoprostol or other prostaglandins; contraindication to vaginal birth Age (mean ± SD): Misoprostol, 27.6 ± 5.1; PGE_2, 27.4 ± 5.5 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 286.4 ± 7.8 days; PGE_2, 285.5 ± 8.8 days Dating criteria: NR Parity (mean ± SD): Misoprostol, 0.5 ± 0.8; PGE_2, 0.6 ± 0.9 Bishop score (median, with 25% and 75% quartiles): Misoprostol, 4 (2, 5); PGE_2, 4 (2, 6) Other: Indications for induction were as follows: Postterm: 78% Hypertension: 8% Oligohydramnios: 7%	1) Median Apgar scores 2) Apgar score < 7 at 1 minute 3) Apgar score < 7 at 5 minutes 4) Cord pH 5) Base deficit 6) Birthweight 7) Episiotomy 8) Laceration 9) 3rd- or 4th-degree laceration 10) Intact perineum 11) Time from induction to delivery 12) Vacuum-assisted deliveries 13) C-sections	1) Median Apgar scores (with 25% and 75 % quartiles): At 1 minute: Misoprostol: 9 (7, 9) PGE_2: 9 (8, 9) p = 0.67 At 5 minutes: Misoprostol: 9 (8, 9) PGE_2: 9 (9, 10) p = 0.72 2) Apgar score < 7 at 1 minute: Misoprostol: 17/111 (15%) PGE_2: 13/111 (12%) p = 0.43 3) Apgar score < 7 at 5 minutes: Misoprostol: 2/111 (2%) PGE_2: 1/111 (1%) p = 1.00 4) Cord pH (mean ± SD): Misoprostol: 7.28 ± 0.09 PGE_2: 7.28 ± 0.10 p = 0.90 5) Base deficit (mean ± SD): Misoprostol: 5.1 ± 4.0 PGE_2: 5.6 ± 4.5 p = 0.38 6) Birthweight (mean ± SD): Misoprostol: 3728 ± 509 g PGE_2: 3631 ± 493 g (no p-value reported) 7) Episiotomy: Misoprostol: 33/111 (30%) PGE_2: 47/111 (42%) (no p-value reported) RR = 0.72 (95% CI, 0.51-1.02)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (78% of total study population). Results not stratified by parity. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Type(s) of providers: Unspecified hospital Length of follow-up: None	Other: 7%		8) Laceration: Misoprostol: 55/111 (50%) PGE_2: 49/111 (44%) (no p-value reported) RR = 1.16 (95% CI, 0.89-1.51) 9) 3^{rd}- or 4^{th}-degree laceration: Misoprostol: 6/111 (5%) PGE_2: 4/111 (4%) (no p-value reported) RR = 1.55 (95% CI, 0.45-5.31) 10) Intact perineum: Misoprostol: 17/111 (15%) PGE_2: 18/111 (16%) (no p-value reported) RR = 0.97 (95% CI, 0.53-1.78) 11) Time from induction to delivery (mean $\pm$ SD): Misoprostol: 753 $\pm$ 588 minutes PGE_2: 941 $\pm$ 506 minutes p = 0.018 12) Vacuum-assisted deliveries: Misoprostol: 3/111 (3%) PGE_2: 15/111 (14%) (no p-value reported) RR = 0.20 (95% CI, 0.06-0.67) 13) C-sections: Misoprostol: 15/111 (14%) PGE_2: 12/111 (11%) (no p-value reported) RR = 1.25 (95% CI, 0.61-2.55)	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units, 1994	Design: RCT, randomization by computer-generated random numbers Interventions: 1) PGE_2 gel + induction by oxytocin (n = 174) Protocol: PGE_2 gel (0.5 mg) inserted into intracervical canal within 24 hours of randomization. No repeat applications. FHR and uterine contractions monitored continuously for ≥ 4 hours. If no labor after 12 hours, then patient induced using amniotomy (where clinically feasible), followed by oxytocin infusion ("according to a uniform protocol"). If no active labor 24 hours after oxytocin infusion, then C-section performed or induction of labor continued. (Decision to perform C-section not dictated by study protocol.) 2) Placebo gel + induction by oxytocin (n = 91) Protocol: Same as in 1), above, except that placebo gel used instead of PGE_2 gel. 3) Monitoring (n = 175) Protocol: Weekly cervical exam + twice-weekly NST and U/S assessment of AFV. Spontaneous labor awaited, but labor could be induced if: Bishop score > 6; estimated fetal weight > 4500 g; medical or obstetric indication for delivery developed; largest pocket of amniotic fluid < 2	No. of subjects at start: 440 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 440 Inclusion criteria: Gestational age ≥ 287 days and < 301 days Exclusion criteria: Medical or obstetric complications requiring induction, C-section, or frequent monitoring; estimated fetal weight > 4500 g; Bishop score ≥ 7; non-reactive NST; amniotic fluid pocket < 2 cm Age (mean ± SD): PGE_2-oxytocin: 25.4 ± 5.7 Placebo-oxytocin: 25.4 ± 5.3 Monitoring: 26.1 ± 5.8 Race: PGE_2-oxytocin: 67% White, 32% Black, 1% not available Placebo-oxytocin: 63% White, 37% Black Monitoring: 60% White, 38% Black, 2% not available Gestational age at entry: PGE_2-oxytocin: 8l% 287-293 days; 19% 295-301 days Placebo-oxytocin: 79% 287-293 days; 21% 295-301 days Monitoring: 79% 287-293 days; 21% 295-301 days Dating criteria: Any one of following: 1) LMP + audible fetal heartbeat documented for ≥ 21 weeks by fetoscope or ≥ 30	1) Mechanical ventilation 2) Meconium aspiration 3) Nerve injury 4) Seizures 5) ≥ 1 adverse neonatal outcome 6) Apgar score < 4 at 5 minutes 7) Birthweight (mean) 8) Birthweight ≥ 4500 g 9) Time from randomization to delivery 10) Gestational age at delivery 11) Maternal infection 12) Maternal transfusion 13) Hyperstimulation 14) C-sections	1) Mechanical ventilation: PGE_2-oxytocin: 0 Placebo-oxytocin: 1/91 (1%) Monitoring: 1/175 (< 1%) (no p-value reported) 2) Meconium aspiration: PGE_2-oxytocin: 1/174 (< 1%) Placebo-oxytocin: 1/91 (1%) Monitoring: 2/175 (1%) (no p-value reported) 3) Nerve injury: PGE_2-oxytocin: 1/174 (< 1%) Placebo-oxytocin: 0 Monitoring: 0 (no p-value reported) 4) Seizures: PGE_2-oxytocin: 0 Placebo-oxytocin: 2/91 (2%) Monitoring: 1/175 (< 1%) (no p-value reported) 5) ≥ 1 adverse neonatal outcome: PGE_2-oxytocin: 1/174 (< 1%) Placebo-oxytocin: 3/91 (3%) Monitoring: 1/175 (< 1%) (no p-value reported) 6) Apgar score < 4 at 5 minutes: PGE_2-oxytocin: 0 Placebo-oxytocin: 0 Monitoring: 1/175 (< 1%) (no p-value reported) 7) Birthweight (mean ± SD): PGE_2-oxytocin: 3607 ± 382 g Placebo-oxytocin: 3532 ± 464 g Monitoring: 3606 ± 440 g (no p-value reported)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Sample size estimates based on perinatal morbidity/mortality and maternal mortality.

(continued on next page)

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	cm; or abnormal NST followed by positive CST. If NST nonreactive, but CST negative, then testing repeated in 24 hours. Patients undelivered by 308 days (44 completed weeks) were released from the protocol and managed as "appropriate for the clinical situation." Dates: Dec 1987 - July 1989 Location: Multiple sites in US Setting: University hospitals Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	weeks by Doppler; 2) LMP + compatible uterine size estimation at ≤ 24 weeks; 3) LMP + positive pregnancy test obtained early enough to assure that gestation exceeded 41 weeks; 4) if LMP uncertain, then fetal heartbeat documented for ≥ 32 weeks by Doppler; 5) U/S before 26 weeks Parity (% nulliparous): PGE_2-oxytocin: 60% Placebo-oxytocin: 59% Monitoring: 54% Bishop score (mean ± SD): PGE_2-oxytocin: 4.0 ± 1.4 Placebo-oxytocin: 3.8 ± 1.4 Monitoring: 3.9 ± 1.5		8) Birthweight ≥ 4500 g: PGE_2-oxytocin: 1/174 (< 1%) Placebo-oxytocin: 3/91 (3%) Monitoring: 6/175 (4%) (no p-value reported) 9) Time from randomization to delivery (median, with range): PGE_2-oxytocin: 36 hours (6-492) Placebo-oxytocin: 35 hours (7-487) Monitoring: 85 hours (5-538) p < 0.001 10) Gestational age at delivery: 287-293 days / 294-301 days / >302 days PGE_2-oxy: 64% / 34% / 1% Placebo-oxy: 66% / 32% / 2% Monitoring: 38% / 47% / 14% p < 0.001 11) Maternal infection: PGE_2-oxytocin: 33/174 (19%) Placebo-oxytocin: 13/91 (14%) Monitoring: 25/175 (14%) p = not significant 12) Maternal transfusion: PGE_2-oxytocin: 2/174 (1%) Placebo-oxytocin: 0 Monitoring: 3/175 (2%) p = not significant 13) Hyperstimulation: PGE_2-oxytocin: 2/174 (1%) Placebo-oxytocin: 1/91 (1%) Monitoring: 0 p = not significant 14) C-sections: PGE_2-oxytocin: 39/174 (22%) Placebo-oxytocin: 16/91 (18%) Monitoring: 32/175 (18%) p = not significant	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
O'Brien, Mercer, Cleary, et al., 1995	Design: RCT, randomization by table of random numbers Interventions: 1) PGE_2 gel (n = 50) Protocol: PGE_2 (2 mg) gel given intravaginally every day for 5 consecutive days. Patients monitored for minimum of 30 minutes after each dose. At 41 weeks, patients re-evaluated. If cervix favorable, NST non-reactive with a BPS ≤ 6, oligohydramnios, FHR decelerations, or evidence of growth restriction, then patient induced. Otherwise, patients evaluated with twice –weekly NSTs and weekly AFV assessments. 2) Placebo gel (n = 50) Protocol: Same as above, except that placebo gel used instead of PGE_2. Dates: June 1993 - June 1994 Location: Memphis, TN Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 100 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 100 Inclusion criteria: 38-40 weeks gestation; Bishop score ≤ 6; no medical indication for delivery; ≤ 1 previous low-transverse C-section Exclusion criteria: Nonreactive NST; oligohydramnios (AFI < 5.0 cm); macrosomia (estimated fetal weight > 4000 g); fetal growth restriction (estimated fetal weight < 10^{th} percentile) Age: NR Race: NR Gestational age at entry (mean ± SD): PGE_2, 38.9 ± 0.54 weeks; placebo, 39.0 ± 0.66 weeks Dating criteria: NR Parity: PGE_2, 40% nulliparous; placebo, 56% nulliparous Bishop score (median, with range): PGE_2, 4 (1-6); placebo, 4 (1-6)	1) Birthweight 2) Macrosomia 3) Apgar score < 7 at 5 minutes 4) Admission to NICU 5) Meconium staining 6) Postdate pregnancies (delivery > estimated date) 7) Postterm pregnancies (delivery ≥ 294 days) 8) Inpatient inductions 9) Gestational age at delivery 10) Chorioamnionitis 11) C-sections	1) Birthweight (mean ± SD): PGE_2: 3320 ± 400 g Placebo: 3450 ± 400 g p = 0.11 2) Macrosomia: PGE_2: 1/50 (2%) Placebo: 4/50 (8%) p = 0.36 3) Apgar score < 7 at 5 minutes: PGE_2: 0 Placebo: 2/50 (4%) p = 0.50 4) Admission to NICU: PGE_2: 1/50 (2%) Placebo: 5/50 (10%) p = 0.20 5) Meconium staining: PGE_2: 8/50 (16%) Placebo: 15/50 (30%) p = 0.15 6) Postdate pregnancies (delivery > estimated date): PGE_2: 20/50 (40%) Placebo: 33/50 (66%) p =0.016 7) Postterm pregnancies (delivery ≥ 294 days): PGE_2: 2/50 (4%) Placebo: 3/50 (6%) p = 1.0 8) Inpatient inductions: PGE_2: 6/50 (12%) Placebo: 14/50 (28%) p = 0.08	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not stratified by parity.

(continued on next page)

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				9) Gestational age at delivery (mean ± SD): PGE_2: 39.9 ± 1.0 weeks Placebo: 40.5 ± 0.99 weeks p = 0.003 10) Chorioamnionitis: PGE_2: 4/50 (8%) Placebo: 7/50 (14%) p = 0.52 11) C-sections: PGE_2: 7/50 (14%) Placebo: 10/50 (20%) p = 0.59	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Papageorgiou, Tsionou, Minaretzis, et al., 1992	Design: RCT, allocation to treatment group by even/odd admission number Interventions: 1) PGE_2 gel (n = 83) Protocol: PGE_2 gel (0.5 mg) instilled deeply into cervical canal by syringe. Patient monitored for 45 min before and after treatment. Pelvic exam done 6 hours after placement of gel. If Bishop score < 5, then second dose given. Pelvic exam repeated 6 hours after second dose. If Bishop score still < 5, then patient considered to have failed PGE_2 ripening and given oxytocin infusion. If Bishop score > 5, but regular contractions or progressive dilatation not observed, then oxytocin used for labor augmentation. 2) Oxytocin (n = 82) Protocol: Up to 3 trials of oxytocin infusion, each lasting 4 hours, with 4-hour rest period between trials. Infusion started at 5 mU/min and increased by 5 mU/min every half hour up to 30 mU/min. If no labor established after 3 trials, then patient delivered by C-section. Dates: NR Location: Athens, Greece Setting: University hospital	No. of subjects at start: 165 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 165 Inclusion criteria: Singleton pregnancy; vertex presentation; unripe cervix; no other obstetric complications; 41 completed weeks' gestation; nonreactive NST; normal AFI by U/S Exclusion criteria: None specified Age (mean ± SEM): PGE_2, 24.9 ± 0.5; oxytocin, 25.0 ± 0.5 Race: NR Gestational age at entry: NR (required to have completed 41 weeks' gestation for entry into study) Dating criteria: LMP + U/S at 20 weeks Parity (mean ± SEM): PGE_2, 1.6 ± 0.1; oxytocin, 1.5 ± 0.1 Bishop score (mean ± SEM): PGE_2, 2.9 ± 0.1; oxytocin, 3.1 ± 0.1	1) Apgar score < 7 at 5 minutes 2) Birthweight 3) C-sections for disproportion 4) C-sections for fetal distress 5) Vacuum delivery 6) Vaginal delivery 7) Hyperstimulation	1) Apgar score < 7 at 5 minutes: PGE_2: 2/83 (2.4%) Oxytocin: 8/82 (9.7%) p < 0.05 2) Birthweight (mean ± SEM): PGE_2: 3601 ± 55 g Oxytocin: 3562 ± 43 g p = not significant 3) C-sections for disproportion: PGE_2: 4/83 (4.8%) Oxytocin: 4/82 (4.8%) p = not significant 4) C-sections for fetal distress: PGE_2: 2/83 (2.4%) Oxytocin: 3/82 (3.6%) p = not significant 5) Vacuum delivery: PGE_2: 7/83 (8.4%) Oxytocin: 9/82 (10.9%) p = not significant 6) Vaginal delivery: PGE_2: 74/83 (89%) Oxytocin: 58/82 (70.7%) p < 0.01 7) Hyperstimulation: PGE_2: 2/83 (2.4%) Oxytocin: 4/82 (4.8%) p = not significant	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: - Statistical tests: + Gestational age: - Dating criteria: + Bishop score: + Results not stratified by parity.

(continued on next page)

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Type(s) of providers: Unspecified OB/GYN Length of follow-up: None				

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Rayburn, Gosen, Ramadei, et al., 1988	Design: RCT, randomization by drawing a card Interventions: 1) Prostaglandin E_2 (PGE_2) gel (2.5 mg) (n = 55) 2) Placebo gel (n = 63) Treatment protocol: After assignment of Bishop score and a reactive NST, gel instilled into cervix using a 16-gauge angiocatheter tube. Patient remained in semi-Trendelenburg position while uterine contractions and FHR monitored for 2 hours. Induction of labor with oxytocin scheduled approximately 12 hours after instillation of study drug. Induction followed ACOG guidelines. Dates: Dec 1985 - Feb 1987 Location: Omaha, NE Setting: University hospital and military hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 118 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 118 Inclusion criteria: Singleton pregnancy; scheduled for induction at 42 weeks; unfavorable cervix (Bishop score ≤ 5) Exclusion criteria: None specified Age (mean ± SD, with range): PGE_2: 23 ± 1.2 (21.8 to 24.2) Placebo: 24 ± 1.6 (22.4 to 25.6) Race: NR Gestational age at entry: 42 weeks Dating criteria: LMP *plus* "compatible clinical milestones" or U/S results from first half of gestation Parity: PGE_2: 51% nulliparous Placebo: 63% nulliparous Bishop score (mean ± SD): PGE_2: 3.2 ± 1.0 Placebo: 3.4 ± 0.8	1) Vaginal delivery, spontaneous 2) Vaginal delivery, forceps-assisted 3) C-sections 4) Time to delivery	1) Vaginal delivery, spontaneous: PGE_2: 42/55 (76%) Placebo: 35/63 (56%) p < 0.05 2) Vaginal delivery, forceps-assisted: PGE_2: 3/55 (5.5%) Placebo: 7/63 (11%) p < 0.05 3) C-sections: *Overall:* PGE_2: 10/55 (18%) Placebo: 21/63 (33%) p < 0.05 *For fetal distress:* PGE_2: 1/55 (2%) Placebo: 6/63 (9.5%) (no p-value reported) *For failure to progress:* PGE_2: 9/55 (16%) Placebo: 13/63 (21%) (no p-value reported) *For other reasons:* PGE_2: 0 Placebo: 2/63 (3%) (no p-value reported) 4) Time to delivery (mean ± SD): PGE_2: 5.5 ± 1.6 hours Placebo: 9.5 ± 2.3 hours p < 0.01	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: - Statistical tests: - Gestational age: + Dating criteria: + Bishop score: +

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Sala-malekis, Vitoratos, Kassanos, et al., 2000	Design: RCT, method of randomization not described Interventions: 1) Membrane stripping (n = 34) Protocol: Examiner's finger inserted as far as possible through the internal cervical os, separating the membranes from the lower uterine segment and rotating 360º. Patients followed up for 4 days. 2) Oxytocin (n = 35) Protocol: Oxytocin infusion given over 6 hours. Initial infusion 0.5 mU/min, then doubled hourly, reaching a maximum of 4 mU/min. Continuous cardiotocographic monitoring throughout 6-hour infusion period. Patients followed up for 4 days. 3) Vaginal exam (control) (n = 35) Protocol: "Gentle vaginal examination" given. Patients followed up for 4 days. Dates: NR Location: Athens, Greece Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 104 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 104 Inclusion criteria: Primigravida; gestational age 40-41 weeks; Bishop score ≤ 5; no maternal or fetal complications; singleton pregnancy; cephalic presentation Exclusion criteria: None specified Age (mean ± SD): Stripping: 26 ± 2.4 Oxytocin: 27.1 ± 4.5 Control: 26.3 ± 3.8 Race: 100% Greek Gestational age at entry (mean ± SD): Stripping: 283.3 ± 2.4 days Oxytocin: 284.1 ± 2.1 days Control: 282.9 ± 3.2 days Dating criteria: Clinical exam and U/S during 1st trimester Parity: 100% primigravida Bishop score: NR (required to be ≤ 5 for entry into study)	1) C-sections 2) Chorioamnionitis 3) Inductions 4) Spontaneous labor 5) Time to onset of labor	1) C-sections: Stripping: 2/34 (5.9%) Oxytocin: 3/35 (8.6%) Control: 1/35 (2.9%) p = not signifcant 2) Chorioamnionitis: Stripping: 0 Oxytocin: 0 Control: 0 p = not significant 3) Inductions: Stripping: 1/34 (2.9%) Oxytocin: 2/35 (5.7%) Control: 7/35 (20%) p = 0.05 4) Spontaneous labor: Stripping: 23/34 (67.6%) Oxytocin: 18/35 (51.4%) Control: 12/35 (34.2%) p = 0.05 5) Time to onset of labor (mean ± SD): Stripping: 1.9 ± 1.2 days Oxytocin: 2.1 ± 0.8 days Control: 2.5 ± 0.9 days p = not significant	QUALITY SCORE: Randomized: + Method of randomization: - Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Definition of "labor" used not reported.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Sanchez-Ramos, Kaunitz, Del Valle, et al., 1993	Design: RCT, randomization by table of random numbers (generated by consecutive coin toss) and sealed envelopes Interventions: 1) Misoprostol (n = 64) Protocol: 50-µg misoprostol tablet placed in posterior vaginal fornix. Dose repeated every 4 hours until adequate labor achieved (3 contractions in 10 minutes). Maximum dose = 600 µg. Artificial rupture of the membranes performed as soon as cervical dilatation permitted. Patients in active labor with arrest of dilatation (no change in dilatation for 2+ hours at 5 cm or more) received oxytocin augmentation. 2) Oxytocin (n = 65) Protocol: Oxytocin infusion started at 1-2 mU/minute and gradually increased in dose increments of 1-2 mU/minute at 30-min intervals, as needed. If Bishop score < 5 before start of oxytocin infusion, then cervical ripening was performed with single or multiple doses of PGE_2 gel. Dates: Jan-Aug 1992 Location: Jacksonville, FL Setting: University hospital Type(s) of providers: Unspecified OB/GYN	No. of subjects at start: 130 Dropouts: 1 (excluded after randomization for breech presentation) Loss to follow-up: NA No. of subjects at end: 129 Inclusion criteria: Obstetric indication for labor; medical complications (including diabetes and renal disease); absence of labor or fetal distress; no previous C-section or other uterine surgery; singleton pregnancy with vertex presentation; no contraindications to vaginal delivery Exclusion criteria: None specified Age (mean + SD): Misoprostol, 23.7 ± 5.5; oxytocin, 23.1 ± 5.6 Race: Misoprostol, 50% non-White; oxytocin, 51% non-White Gestational age at entry (mean ± SD): Misoprostol, 38.8 ± 2.6 weeks; oxytocin, 38.8 ± 4.0 weeks Dating criteria: NR Parity (mean ± SD): Misoprostol, 0.8 ± 1.2; oxytocin, 0.7 ± 1.1 Bishop score (mean ± SD): Misoprostol, 4.0 ± 2.2; oxytocin, 4.2 ± 2.2 Other: Indications for induction were as follows: Preeclampsia: 34%	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Cord pH < 7.16 5) Admission to NICU 6) Bleeding > 500 ml 7) Forceps delivery 8) Vacuum delivery 9) C-sections 10) Induction-agent costs 11) Time to delivery	1) Apgar score < 7 at 1 minute: Misoprostol: 11/64 (17.2%) Oxytocin: 9/65 (13.8%) p = not significant 2) Apgar score < 7 at 5 minutes: Misoprostol: 1/64 (1.6%) Oxytocin: 1/65 (1.5%) p = not significant 3) Birthweight (mean ± SD): Misoprostol: 3181.5 ± 731.8 g Oxytocin: 3231.4 ± 662.8 g p = not significant 4) Cord pH < 7.16: Misoprostol: 9/64 (14.1%) Oxytocin: 7/65 (10.8%) p = not significant 5) Admission to NICU: Misoprostol: 3/64 (4.7%) Oxytocin: 6/65 (9.2%) p = not significant 6) Bleeding > 500 ml: Misoprostol: 1/64 (1.6%) Oxytocin: 0/65 p = not significant 7) Forceps delivery: Misoprostol: 9/64 (14.1%) Oxytocin: 9/65 (13.8%) p = not significant 8) Vacuum delivery: Misoprostol: 4/64 (6.3%) Oxytocin: 7/65 (10.8%) p = not significant 9) C-sections: Misoprostol: 14/64 (21.9%) Oxytocin: 14/65 (21.5%) p = not significant	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (19% of total study population). Results not stratified by parity. Study underpowered to detect differences in some outcomes (e.g., hyperstimulation 11% in misoprostol group, 4.6% in oxytocin group, but not significant). Sample size estimates based on time to delivery. Total dose and maximum rate of oxytocin significantly lower in misoprostol group. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Length of follow-up: None	Postterm: 19% (22% misoprostol, 15% oxytocin) PROM: 13% Abnormal fetal testing: 9% Diabetes: 7% Other: 18%		10) Induction-agent costs (per patient): Misoprostol (± oxytocin): $49 Oxytocin alone: $205 Oxytocin + PGE_2: $315 (no p-value reported) 11) Time to delivery (mean ± SD): Misoprostol: 661.9 ± 435.9 minutes Oxytocin: 1104.9 ± 968.1 minutes p = 0.004	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Satin, Hankins, and Yeomans, 1991	Design: RCT, randomization by sealed envelope Interventions: 1) Oxytocin, slow dose escalation (n = 32) Protocol: Initial dose 2 mU/min. Incremental increases of 1 mU/min given at 30-minute intervals to maximum dose of 40 mU/min. 2) Oxytocin, fast dose escalation (n = 48) Protocol: : Initial dose 2 mU/min. Incremental increases of 2 mU/min given at 15-minute intervals to maximum dose of 40 mU/min. In both groups, oxytocin doses were increased until an adequate labor pattern was achieved (defined as labor resulting in cervical change). Amniotomy performed in active labor. Internal FHR and pressure monitored. Pressure catheter used to titrate. Induction considered to have failed if no cervical dilatation or spontaneous rupture of membranes by 8-10 hours and no evidence of fetal distress or maternal illness. Dates: NR Location: San Antonio, TX Setting: Military hospital Type(s) of providers: Unspecified OB/GYN	No. of subjects at start: 80 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 80 Inclusion criteria: Cervical dilatation ≤ 2 cm; Bishop score ≤ 6; no regular uterine activity; intact membranes Exclusion criteria: Malpresentation; placenta previa; active herpes infection; hypertension; deviation from dosing protocol Age (mean + SD): Slow, 24.3 ± 3.6; fast, 24.7 ± 3.2 Race: NR Gestational age at entry: Mean NR. Slow, 31/32 (97%) ≥ 42 weeks; fast, 44/48 (92%) ≥ 42 weeks Dating criteria: NR Parity: Slow, 47% nulliparous; fast, 46% nulliparous Bishop score: NR	1) Apgar score ≤ 3 at 1 minute 2) Apgar score ≤ 6 at 5 minutes 3) Birthweight 4) Use of epidural 5) Induction failure 6) Hyperstimulation/FHR abnormalities requiring oxytocin to be stopped 7) C-sections (by parity) 8) Mid-forceps delivery (by parity) 9) Time to delivery (by parity)	1) Apgar score ≤ 3 at 1 minute: Slow: 0/32 Fast: 1/48 (2%) p = not significant 2) Apgar score ≤ 6 at 5 minutes: Slow: 1/32 (3%) Fast: 1/48 (2%) p =not significant 3) Birthweight (mean ± SD): Slow: 3623 ± 459 g Fast: 3670 ± 516 p = not significant 4) Use of epidural: Slow: 25% Fast: 27% p = not significant 5) Induction failure: Slow: 10/32 (31%) Fast: 4/48 (8%) p < 0.05 6) Hyperstimulation/FHR abnormalities requiring oxytocin to be stopped: Slow: 66%, 0 episodes; 25%, 1 episode; 3%, 2 episodes; 6%, ≥ 3 episodes Fast: 46%, 0 episodes; 29%, 1 episode; 8%, 2 episodes; 17%, ≥ 3 episodes p = not significant 7) C-sections (by parity): Slow, nulliparous: 1/32 (3%) Slow, multiparous: 0 Fast, nulliparous: 3/48 (6%) Fast, multiparous: 2/48 (4%) p = not significant 8) Mid-forceps delivery (by parity): Slow, nulliparous: 1/32 (3%) Slow, multiparous: 1/32 (3%)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: - Hyperstimulation more common in fast protocol, but study underpowered to detect difference *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Length of follow-up: None			Fast, nulliparous: 2/48 (4%) Fast, multiparous: 0 p = not significant 9) Time to delivery (mean, by parity): Slow, nulliparous: 15 hours, 18 minutes Fast, nulliparous: 9 hours, 16 minutes p < 0.05 Slow, multiparous: 10 hours, 54 minutes Fast, multiparous: 8 hours, 2 minutes p < 0.05	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Sawai, O'Brien, Mastro-giannis, et al., 1994	Design: RCT, computer-generated randomization Interventions: 1) Self-administered PGE_2 suppositories (2 mg) (n = 38) Protocol: Patients given explicit instructions on how to avoid intracervical placement of suppository. Enough suppositories given for daily use until next clinic visit. Telephone contact with investigator available on 24-hour basis. Patients returned for weekly sonogram for AFI and twice-weekly NST and Bishop scoring. Suppositories dispense at each clinic visit until spontaneous labor occurred or until patient admitted for induction of labor for Bishop score ≥ 9, oligohydramnios (AFI < 5 cm), "nonreassuring" FHR tracing, gestational age of 44 weeks, or the development of preeclampsia or other exclusion criteria. 2) Placebo suppositories (n = 42) Protocol: Same as above, except that placebo suppositories used instead of PGE_2. Dates: May 1990 - Sep 1991 Location: Tampa, FL Setting: University hospital Type(s) of providers:	No. of subjects at start: 91 Dropouts: 11 Loss to follow-up: NA No. of subjects at end: 80 Inclusion criteria: Gestational age ≥ 41 weeks; uncomplicated pregnancy; Bishop score < 9; reactive NST; normal U/S Exclusion criteria: Maternal medical problems; previous uterine surgery; previous stillbirth; abnormal FHR; vaginal bleeding; spontaneous rupture of membranes; regular uterine contractions; abnormal U/S findings; estimated fetal weight ≥ 4500 g Age: NR Race: NR Gestational age at entry (mean ± SD): PGE_2, 297.0 ± 5.4 days; placebo, 295.0 ± 4.5 days (p = 0.021) Dating criteria: NR ("reliable dating criteria") Parity: NR Bishop score: Baseline scores not reported	1) Apgar score < 7 at 5 minutes 2) Birthweight 3) Umbilical artery pH 4) Admission to NICU 5) C-sections 6) Chorioamnionitis 7) Time from admission to delivery 8) Antepartum testing charges (per patient)	1) Apgar score < 7 at 5 minutes: PGE_2: 1/38 (2.6%) Placebo: 1/42 (2.4%) p = not significant 2) Birthweight (mean ± SD): PGE_2: 3.50 ± 0.40 kg Placebo: 3.68 ± 0.39 kg p = 0.051 3) Umbilical artery pH (mean ± SD): PGE_2: 7.27 ± 0.07 Placebo: 7.27 ± 0.07 p = not significant 4) Admission to NICU: PGE_2: 2/38 (5.3%) Placebo: 4/42 (9.5%) p = not significant 5) C-sections: PGE_2: 1/38 (2.6%) Placebo: 6/42 (14.3%) p = not significant 6) Chorioamnionitis: PGE_2: 2/38 (5.3%) Placebo: 10/42 (24%) p = 0.04 7) Time from admission to delivery (mean ± SD): *Nulliparas:* PGE_2 (n = NR): 10.7 ± 5.1 hours Placebo (n = NR): 15.3 ± 7.6 hours p = 0.035 *Multiparas:* PGE_2 (n = NR): 11.2 ± 1.3 hours Placebo (n = NR): 7.1 ± 4.4 hours p = not significant	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: - Baseline characteristics not reported. Underpowered to detect differences in categorical variables. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Unspecified OB/GYN Length of follow-up: None			8) Antepartum testing charges (per patient; mean ± SD): *All patients:* PGE_2: $476.97 ± $170.36 Placebo: $647.29 ± $257.36 p = 0.001 *Nulliparas:* PGE_2 (n = NR): $456.44 ± $141.55 Placebo (n = NR): $659.67 ± $271.38 p = 0.006 *Multiparas:* PGE_2 (n = NR): $495.45 ± $194.50 Placebo (n = NR): $630 ± $244.12 p = not significant	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Sawai, Williams, O'Brien, et al., 1991	Design: RCT, randomization by sealed envelope Interventions: 1) PGE_2 gel (n = 24; 14 nulliparas and 10 multiparas) Protocol: PGE_2 gel (2 mg) placed in the posterior vaginal fornix. Uterine activity and FHR tracings monitored for 1-2 hours after gel insertion. If no regular uterine contractions and NST reactive, then patient discharged and asked to return for weekly sonograms for AFI assessment and twice-weekly NSTs, cervical scoring, and application of gel. Labor induced if spontaneous labor did not occur and Bishop score > 9, if oligohydramnios present (AFI < 5 cm), if FHR tracing "not reassuring," or if a gestational age of 44 weeks was reached. 2) Placebo gel (n = 26; 16 nulliparas and 10 multiparas) Protocol: Same as above, except that placebo gel used instead of PGE_2. Dates: Aug 1988 - Aug 1989 Location: Tampa, FL Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 50 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 50 Inclusion criteria: Gestational age ≥ 287 days; unfavorable cervix (Bishop score < 9) Exclusion criteria: Diabetes; hypertension; previous uterine surgery; abnormal FHR tracings; vaginal bleeding; spontaneous rupture of membranes; regular uterine contractions; nonvertex presentation; macrosomia (estimated fetal weight > 4500 g); fetal anomalies; fetal growth retardation; oligohydramnios; multiple gestation Age: NR Race: NR Gestational age at entry: NR (gestational age ≥ 287 days required for entry into study) Dating criteria: LMP confirmed by early clinical exam and/or early U/S Parity: PGE_2, 58% nulliparous; placebo, 62% nulliparous Bishop score: NR; score < 9 required for entry into study	1) Apgar scores at 1 minute 2) Apgar scores at 5 minutes 3) Birthweight 4) Umbilical arterial blood pH 5) Admission to NICU 6) C-sections 7) Length of labor and delivery	1) Apgar scores at 1 minute (median): PGE_2 nulliparas (n = 14): 9.0 Placebo nulliparas (n = 16): 8.5 p = not significant PGE_2 multiparas (n = 10): 9.0 Placebo multiparas (n = 10): 9.0 p = not significant 2) Apgar scores at 5 minutes (median): PGE_2 nulliparas: 9.0 Placebo nulliparas: 9.0 p = not significant PGE_2 multiparas: 9.0 Placebo multiparas: 9.0 p = not significant 3) Birthweight (mean ± SEM): PGE_2 nulliparas: 3753.6 ± 126 Placebo nulliparas: 3910.7 ± 113 p = not significant PGE_2 multiparas: 3564.5 ± 119 Placebo multiparas: 3589.0 ± 74 p = not significant 4) Umbilical arterial blood pH (mean ± SEM): PGE_2 nulliparas: 7.28 ± 0.02 Placebo nulliparas: 7.28 ± 0.02 p = not significant PGE_2 multiparas: 7.32 ± 0.01 Placebo multiparas: 7.19 ± 0.06 p = not significant 5) Admission to NICU: PGE_2 nulliparas: 0 Placebo nulliparas: 0 p = not significant PGE_2 multiparas: 0	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: - Study underpowered to detect differences in categorical variables (e.g., C-sections). *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				Placebo multiparas: 2 p = not significant 6) C-sections (all for failure to progress or arrest of descent): PGE_2 nulliparas: 6/14 (43%) Placebo nulliparas: 3/16 (19%) p = not significant PGE_2 multiparas: 0 Placebo multiparas: 1/10 (10%) p = not significant 7) Length of labor and delivery (mean ± SEM): PGE_2 nulliparas: 17.6 ± 2.7 hours Placebo nulliparas: 13.9 ± 1.9 hours p = not significant PGE_2 multiparas: 5.4 ± 2.0 hours Placebo multiparas: 8.2 ± 1.2 hours p = not significant	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Sciscione, Nguyen, Manley, et al., 2001	Design: RCT, randomization by computer-generated list of random numbers and sealed envelopes Interventions: 1) Transcervical Foley catheter (n = 58) Protocol: 16F Foley catheter with 30-ml balloon inserted into endocervical canal under direct visualization via a sterile speculum exam. Effort was made not to touch the catheter to vagina or ectocervix. Once balloon in place, 30 ml water injected. Traction applied by taping end of catheter to patient's leg. Catheter checked for extrusion every 6 hours by cervical exam. If not extruded, then catheter adjusted to maintain traction. FHR monitoring started after placement, and patient allowed to ambulate. Oxytocin given after catheter extrusion, beginning a 1 mIU and increasing 1 mIU every 15 minutes. Artificial rupture of membranes done as soon as clinically feasible. 2) Misoprostol (n = 53) Protocol: 50-µg tablet placed in posterior vaginal fornix every 4 hours to maximum of 6 doses. Dosing suspended in the event of onset of labor, uterine tachysystole, non-reassuring FHR, or rupture of membranes. Oxytocin started (as above) 4 hours after last dose of misoprostol in women	No. of subjects at start: 114 Dropouts: 3 (2 for protocol violations; 1 for failure to meet inclusion criteria) Loss to follow-up: NA No. of subjects at end: 111 Inclusion criteria: Admitted for labor induction; single gestation; vertex presentation; > 28 weeks' gestation; Bishop score < 6 Exclusion criteria: Rupture of membranes; antepartum bleeding; active genital herpes infection; fetal death; placenta previa; previous induction or preinduction agent during pregnancy; known allergy to misoprostol Age (mean ± SD): Catheter, 25.1 ± 6.9; misoprostol, 25.9 ± 6.9 Race: NR Gestational age at entry: NR Dating criteria: NR Parity: Catheter, 70.6% nulliparous; misoprostol, 71.7% nulliparous Bishop score (median): Catheter, 3.0; misoprostol, 2.0 Other: Indications for induction: Preeclampsia: 32% Oligohydramnios: 25% Postterm: 14% Growth restriction: 8%	1) Birthweight 2) C-sections 3) Delivery within 24 hours 4) Vaginal delivery within 24 hours	1) Birthweight (mean ± SD): Catheter: 2979.5 ± 619.9 g Misoprostol: 2969.8 ± 743.7 g p = 0.94 2) C-sections: *Overall:* Catheter: 31.8% Misoprostol: 37.8% p = 0.46 *For nonreassuring FHR tracing::* Catheter: 12% Misoprostol: 24% p = 0.09 3) Delivery within 24 hours: Catheter: 54.5% Misoprostol: 67.9% p = 0.31 4) Vaginal delivery within 24 hours: Catheter: 73% Misoprostol: 84% p = 0.23	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: - Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (14% of total study population). Results not stratified by parity. Sample size estimates based on change in Bishop score. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	not in active labor, but with Bishop scores > 5, or after 6 doses. Artificial rupture of membranes done as soon as clinically feasible. Dates: July 1997 - July 1999 Location: Newark, DE Setting: Community hospital Type(s) of providers: General OB/GYN; residents Length of follow-up: None	Elective: 5% Chronic hypertension: 3% Diabetes: 3% Macrosomia: 3% Other: 8%			

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Srisomboon, Piyamongkol, and Aiewsakul, 1997	Design: RCT, blocked randomization Interventions: 1) Intracervical misoprostol (n = 50) Protocol: 100 µg misoprostol pill crushed in 3 ml sterile jelly. Mixture instilled in endocervical canal with assistance of speculum visualization. 2) Intravaginal misoprostol (n = 50) Protocol: Same mixture as above, but placed in posterior vaginal fornix. Patients in both groups were left in supine position for 1 hour after administration of gel. Vital signs and side effects monitored every 2 hours. Continuous external cardiotocography performed. Patients re-examined at 12 hours. If cervix unfavorable, then 2nd dose of gel given. If cervix became favorable (≥ 6), then amniotomy performed and oxytocin infusion started, if needed. Oxytocin also started if no cervical change occurred or no uterine contractions occurred after 2nd dose. Infusion started at 1-2 mU/min, increased 1-2 mU/min at 30-min intervals. Dates: Aug 1994 - Sep 1995 Location: Chiang Mai, Thailand	No. of subjects at start: 100 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 100 Inclusion criteria: Singleton pregnancy; parity ≤ 3; vertex presentation; obstetric or medical indication for delivery; intact membranes with no prior stripping; Bishop score ≤ 4; gestational age > 35 weeks; no previous C-section or other uterine surgery; no labor or fetal distress; no evidence of cephalopelvic disproportion; no placenta previa, forelying cord, or vasa previa; no contraindication to the use of prostaglandins Exclusion criteria: None specified Age (mean ± SD): Intracervical, 25.8 ± 5.3; intravaginal, 28.1 ± 5.8 Race: NR Gestational age at entry (mean ± SD): Intracervical, 39.7 ± 2.2; intravaginal, 39.2 ± 2.2 Dating criteria: NR Parity (mean ± SD): Intracervical, 1.3 ± 0.5; intravaginal, 1.4 ± 0.5 Bishop score (mean ± SD): Intracervical, 2.6 ± 0.8; intravaginal, 2.6 ± 0.9	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Forceps delivery 5) Vacuum delivery 6) C-sections 7) Post-partum hemorrhage 8) Time to delivery	1) Apgar score < 7 at 1 minute: Intracervical: 3/50 (6%) Intravaginal: 0/50 p = not significant 2) Apgar score < 7 at 5 minutes: Intracervical: 0/50 Intravaginal: 0/50 p = not significant 3) Birthweight (mean ± SD): Intracervical: 2823 ± 426 g Intravaginal: 2833 ± 505 g p = not significant 4) Forceps delivery: Intracervical: 2/50 (4%) Intravaginal: 5/40 (10%) p = not significant 5) Vacuum delivery: Intracervical: 8/50 (16%) Intravaginal: 9/50 (18%) p = not significant 6) C-sections: Intracervical: 3/50 (6%) Intravaginal: 5/50 (10%) p = not significant 7) Post-partum hemorrhage: Intracervical: 0/50 Intravaginal: 1/50 (2%) p = not significant 8) Time to delivery (mean ± SD): Intracervical: 17.0 ± 8.6 hours Intravaginal: 16.4 ± 8.6 hours p = not significant	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: - Statistical tests: - Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (34% of total study population). Results not stratified by parity.

(continued on next page)

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	Other: Indications for induction were as follows: Postterm: 34% (40% intracervical, 28% intravaginal) Pregnancy-induced hypertension: 31% IUGR: 26% Other: 9%			

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Stenlund, Ekman, Aedo, et al., 1999	Design: RCT, randomization by table of random numbers and sealed envelope Interventions: 1) Mifepristone 400 mg (n = 24) Protocol: Bishop score, U/S, and, "in some cases," Doppler performed before starting treatment. Mifepristone 400 mg given as two tablets. If labor did not start, patients returned to hospital at 24 and 48 hours for assessment of Bishop score and FHR monitoring (30 minutes). If Bishop score ≥ 6 at 48 hours and no labor, then labor induced by amniotomy and oxytocin infusion. If Bishop score < 6, then patient given PGE_2 (0.5 mg) intracervically, repeated 12 hours later, if necessary. 2) Placebo (n = 12) Protocol: Same as above, except that identical placebo substituted for mifepristone. Dates: NR Location: Stockholm, Sweden Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 36 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 36 Inclusion criteria: Indication for induction; induction deferrable for 48 hours; Bishop score ≤ 5; single pregnancy in vertex presentation; intact membranes Exclusion criteria: Contraindication to vaginal delivery; oligohydramnios; prior uterine surgery; parity > 4; renal failure; hepatic disorder; adrenal insufficiency; blood-clotting disorder; anticoagulant or corticosteroid therapy during pregnancy Age (mean ± SD): Mifepristone, 27.4 ± 4.6; placebo, 30.3 ± 5.8 Race: NR Gestational age at entry (mean ± SD): Both groups, 295 ± 4 days Dating criteria: U/S performed in week 16 or 17 Parity: Mifepristone, 79% nulliparous; placebo, 58% nulliparous Bishop score (median, with range): Mifepristone, 3 (0 to 5); placebo, 3 (1 to 5)	1) Apgar scores 2) Birthweight 3) Umbilical pH 4) Seizure requiring anticonvulsant treatment 5) Time to onset of labor 6) Percent in labor by 48 hours 7) Labor or ripe cervix within 48 hours 8) Need for PGE_2: 9) C-sections 10) Vacuum extraction 11) Duration of labor	1) Apgar scores: Median Apgar scores were significantly ($p < 0.05$) lower at 1 minute in the mifepristone group, but did not differ between the two treatment groups at 5 or 10 minutes. (Actual scores NR.) 2) Birth weigh (mean ± SD): Mifepristone: 3881 ± 323 g Control: 3779 ± 438 (no p-value reported) 3) Umbilical pH (mean ± SD): Mifepristone (N = 21/24): 7.12 ± 0.15 Control: 7.19 ± 0.09 p = 0.08 4) Seizure requiring anticonvulsant treatment: Mifepristone: 1/24 (4%) Control: 0 (no p-value reported) 5) Time to onset of labor (median, with range): Mifepristone: 24 hrs, 10 min (1 hr, 50 min to 94 hrs, 45 min) Control: 52 hrs (11 hrs, 15 min to 94 hrs, 45 min) (no p-value reported) 6) Percent in labor by 48 hours (with 95% CI): Mifepristone: 81.8% (65.7% to 97.9%) Control: 27.3% (1.0% to 53.6%) $p < 0.05$ 7) Labor or ripe cervix within 48 hours: Mifepristone: 83.3% Control: 41.7% p = 0.008	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Sample size discussed for primary outcome, but not for secondary outcomes Sample size estimates based on proportion of women delivering within 48 hours and on change in Bishop score. Large discrepancy in parity between two groups (more multiparas in mifepristone group). Results not stratified by parity.

(continued on next page)

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				8) Need for PGE_2: Mifepristone: 17% Control: 58% $p < 0.05$ 9) C-sections (all for fetal distress): Mifepristone: 17% Control: 25% p = not significant 10) Vacuum extraction: Mifepristone: 33% Control: 8% p = not significant 11) Duration of labor (median): Mifepristone: 13 hrs, 39 min Control: 8 hrs, 9 min p = not significant	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Surbek, Boesiger, Hoesli, et al., 1997	Design: RCT, randomization performed by pharmacy using random-numbers table Interventions: 1) Misoprostol (n = 50) Protocol: 50-µg misoprostol gelatin capsule placed in posterior vaginal fornix. If adequate contraction pattern not achieved, then further doses given at 6 hours, 24 hours, and 30 hours. Patients not in labor at 48 hours received IV oxytocin. 2) Oxytocin (n = 50) Protocol: Same as above, except that PGE_2 3-mg capsules used instead of misoprostol. Dates: Jan-Nov 1995 Location: Basel, Switzerland Setting: University hospital Type(s) of providers: Unspecified OB/GYN, residents, and midwives Length of follow-up: None	No. of subjects at start: 103 Dropouts: 3 (excluded due to protocol violations) Loss to follow-up: NA No. of subjects at end: 100 Inclusion criteria: Bishop score ≤ 5; reactive stress test; singleton vertex presentation; no labor Exclusion criteria: Fetal mal-presentation; C-section or other prior uterine surgery; contra-indications to prostaglandins Age (mean ± SD): Misoprostol, 28.8 ± 5.4; PGE_2, 30.4 ± 4.7 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 40 ± 1.63 weeks; PGE_2, 40 ± 2.0 Dating criteria: NR Parity: Misoprostol, 60% nulliparous; PGE_2, 50% nulliparous Bishop score (mean ± SD): Misoprostol, 2.4 ± 1.35; PGE_2, 3.0 ± 1.64 Other: Indications for induction: PROM: 37% Postterm: 32% IUGR/oligohydramnios: 14% Hypertensive disorder: 6% Diabetes mellitus: 6% Psychosocial: 5%	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Cord arterial pH 5) Admission to NICU 6) Vaginal operative delivery 7) C-sections	1) Apgar score < 7 at 1 minute: Misoprostol: 4/50 (8%) PGE_2: 6/50 (12%) p = not significant 2) Apgar score < 7 at 5 minutes: Misoprostol: 0/50 PGE_2: 0/50 p = not significant 3) Birthweight (mean ± SD): Misoprostol: 3360 ± 602 g PGE_2: 3419 ± 659 g p = not significant 4) Cord arterial pH (mean ± SD): Misoprostol: 7.25 ± 0.09 PGE_2: 7.23 ± 0.09 p = not significant 5) Admission to NICU: Misoprostol: 0/50 PGE_2: 3/50 (6%) p = not significant 6) Vaginal operative delivery: Misoprostol: 10/50 (20%) PGE_2: 6/50 (12%) p = not significant 7) C-sections: Misoprostol: 6/50 (12%) PGE_2: 7/50 (14%) p = not significant	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (32% of total study population). Results not stratified by parity. Tachysystole less common in PGE_2 group (8% vs. 14%), but difference not significant. Sample size estimates based on proportion of patients delivering within 24 hours.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Toppozada, Anwar, Hassan, et al., 1997	Design: RCT, randomization by computer-generated table of random numbers Interventions: 1) Vaginal misoprostol (n = 20) Protocol: 100-µg tablet applied intravaginally. If positive response (3 contractions/10 minutes, each lasting 45 seconds and inducing changes in the Bishop score), then dose repeated every 3 hours until cervix ≥ 5 cm. If no response to first dose, then 100-µg dose repeated at 3 hours, and 200-µg dose given every 3 hours thereafter until positive response achieved (up to max of 1000 µg). 2) Oral misoprostol (n = 20) Protocol: Same as above, except that tablets administered orally and second dose (rather than third) doubled if no response to first. In both groups, AROM performed and oxytocin given when cervix ≥ 5 cm. Dates: NR Location: Alexandria, Egypt Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 40 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 40 Inclusion criteria: Indication for induction (diabetes, pregnancy-induced hypertension, or postdates); gestational age 37-42 weeks; single viable pregnancy; vertex presentation; Bishop score ≤ 4 Exclusion criteria: Contraindication to induction or prostaglandins Age (mean ± SD): Vaginal, 27.5 ± 4.51; oral, 29.15 ± 5.40 Race: NR Gestational age at entry (mean ± SD): Vaginal, 40.30 ± 1.87 weeks; oral, 40.85 ± 1.57 weeks Dating criteria: NR Parity (mean ± SD): Vaginal, 0.80 ± 0.95; oral, 1.25 ± 1.16 Bishop score (mean ± SD): Vaginal, 2.25 ± 1.69; oral, 1.85 ± 1.39 Other: Indications for induction were diabetes, pregnancy-induced hypertension, or postdates. Proportion of patients in each category not reported.	1) Forceps deliveries 2) Vacuum deliveries 3) C-sections	1) Forceps deliveries: Vaginal: 1/20 (5%) Oral: 0/20 (no p-value reported) 2) Vacuum deliveries: Vaginal: 3/20 (15%) Oral: 2/20 (10%) (no p-value reported) 3) C-sections: Vaginal: 2/20 (10%) Oral: 4/20 (20%) (no p-value reported)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: - Mode of delivery: + Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Proportion of patients who were induced for postterm pregnancy not reported. No separate results reported for this subgroup. Significantly higher incidence of uterine activity and FHR tracing abnormalities in vaginal group.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Varaklis, Gumina, and Stubblefield, 1995	Design: RCT, randomization by table of random numbers and sealed envelopes Interventions: 1) Misoprostol (n = 36) Protocol: 25 µg given intravaginally every 2 hours for a maximum of 6 doses or until patient experience 3 contractions per 10 minutes. 2) PGE_2 gel (n = 33) Protocol: 0.5 mg placed intracervically. Second dose given after 6 hours if patient not having 3 contractions per 10 minutes. In both groups, no further agents were administered once contraction rate reached 3 per 10 minutes. Oxytocin started 12 hours after first dose of induction agent if patient not in active labor. AROM performed at 3 cm. Dates: NR Location: Portland, ME Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 80 Dropouts: 11 Loss to follow-up: NA No. of subjects at end: 6 Inclusion criteria: Medical indication for induction Exclusion criteria: Severe oligohydramnios; nonreactive stress test; prior uterine surgery; malpresentation; multiple gestation; > 3 contractions per 10 minutes; Bishop score > 5 Age (mean ± SD): Misoprostol, 26.75 ± 5.95; PGE_2, 38.96 ± 1.89 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 39.52 ± 2.4 weeks; PGE_2, 38.96 ± 1.89 weeks Dating criteria: Last menstrual period Parity (mean ± SD): Misoprostol, 0.44 ± 0.70; PGE_2, 0.67 ± 1.34 Bishop score: Median, 3 in both groups Other: Reasons for induction not described in detail. Investigators stated that “the reasons for induction, most frequently prolonged pregnancy, were similar in both groups.”	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Cord arterial pH 5) Assisted vaginal deliveries 6) C-sections 7) Time to vaginal delivery	1) Apgar score < 7 at 1 minute: Misoprostol: 7/36 (19%) PGE_2 gel: 7/33 (21%) p = 0.855 2) Apgar score < 7 at 5 minutes: Misoprostol: 1/36 (3%) PGE_2 gel: 1/33 (3%) p = 1.000 3) Birthweight (mean ± SD): Misoprostol: 3.2 ± 0.84 kg PGE_2 gel: 3.33 ± 0.72 kg p = 0.505 4) Cord arterial pH (mean ± SD): Misoprostol: 7.31 ± 0.05 PGE_2 gel: 7.30 ± 0.08 p = 0.632 5) Assisted vaginal deliveries: Misoprostol: 6/36 (17%) PGE_2 gel: 11/33 (33.3%) (no p-value reported) 6) C-sections: Misoprostol: 8/36 (22%) PGE_2 gel: 3/33 (9%) (no p-value reported) 7) Time to vaginal delivery (mean ± SD): Misoprostol: 15.7 ± 8.1 hours PGE_2 gel: 20.7 ± 8.1 hours p = 0.023	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Proportion of patients who were induced for postterm pregnancy not reported. No separate results reported for this subgroup. Results not stratified by parity. Study underpowered to detect differences in categorical outcomes.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Voss, Cumminsky, Cook, et al., 1996	Design: RCT, randomization by computer-generated random number tables Interventions: 1) PGE_2 gel (0.125 mg) (n = 79) Protocol: FHR and contractions monitored for 30 min before treatment, and Bishop score assessed. Gel (2 ml) inserted into cervix at level of internal cervical os. Monitoring continued for 4 hours after insertion. If no labor and Bishop score ≤ 6 at end of 4-hour monitoring period, then second dose of gel instilled, followed by 4 more hours of monitoring. Subsequent management of labor by attending physician and resident staff. 2) PGE_2 gel (0.25 mg) (n = 70) Protocol: Same as above, but with 0.25-mg dosage. 3) PGE_2 gel (0.5 mg) (n = 80) Protocol: Same as above, but with 0.5-mg dosage. Dates: July 1991 - May 1993 Location: Louisville, KY Setting: University hospital and community hospital (2 sites) Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 291 Dropouts: 62 (excluded due to protocol violations) Loss to follow-up: NA No. of subjects at end: 229 Inclusion criteria: Bishop score ≤ 4; induction required Exclusion criteria: Noncephalic presentation; previous vertical C-section; heavy vaginal bleeding; placenta previa; spontaneous labor; abnormal FHR tracing; maternal asthma or glaucoma; history of hypersensitivity to prostaglandin Age (mean, with 95% CI): 0.125 mg: 25.3 (24.1 to 26.6) 0.25 mg: 25.4 (23.9 to 27.0) 0.5 mg: 26.2 (24.6 to 27.8) Race: NR Gestational age at entry (mean, with 95% CI): 0.125 mg: 39.3 weeks (38.8 to 39.9); 29/79 (37%) “postdates” 0.25 mg: 38.5 weeks (37.3 to 39.6); 21/70 (30%) “postdates” 0.5 mg: 39.4 weeks (38.8 to 40.0); 21/80 (26%) Dating criteria: NR Parity: 0.125 mg: 61% nulliparous 0.25 mg: 60% nulliparous 0.5 mg: 69% nulliparous Bishop score: NR	1) FHR abnormality 2) C-sections 3) Change in Bishop score 4) Hyperstimulation 5) Time to a) active phase of labor, b) complete dilatation, and c) delivery (survival analysis)	1) FHR abnormality: 0.125 mg: 21.8% 0.25 mg: 29.9% 0.5 mg: 24.7% p = not significant 2) C-sections: 0.125 mg: 40.8% 0.25 mg: 40.8% 0.5 mg: 36.8% p = not significant 3) Change in Bishop score (mean): 0.125 mg: 2.08 0.25 mg: 1.43 0.5 mg: 1.94 p = not significant 4) Hyperstimulation: 0.125 mg: 7.7% 0.25 mg: 11.9% 0.5 mg: 10.4% p = not significant 5) Time to a) active phase of labor, b) complete dilatation, and c) delivery: Survival analysis showed no significant differences among the three groups for these outcomes.	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: - Sample size: - Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not stratified by parity.

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Wing, Fassett, and Mishell, 2000	Design: RCT, randomization by computer-generated random number sequence and sealed envelopes Interventions: 1) Mifepristone (n = 97) Protocol: Mifepristone 200 mg given by mouth. Patient re-examined in 24 hours. If Bishop score ≥ 7, then labor induced using oxytocin. If Bishop score < 7, FHR tracing reactive, and no contractions, then patient given 25 µg misoprostol intravaginally. Misoprostol repeated every 4 hours until adequate labor established or 24 hours elapsed (maximum 6 doses or 150 µg). Oxytocin used if no active labor after maximum misoprostol dose and for failure to progress in active phase of labor. Oxytocin infused by pump at an initial dose of 1 mU/minute, with incremental increases every 30 minutes to a maximum dose of 22 mU/minute. 2) Placebo (n = 83) Protocol: Same as above, but with placebo rather than mifepristone Dates: Mar 1997 - Jan 1999 Location: Los Angeles, CA Setting: University hospital (2 sites) Type(s) of providers:	No. of subjects at start: 180 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 180 Inclusion criteria: Singleton pregnancy; vertex presentation; reactive NST; intact membranes; gestational age > 41 weeks; maternal age > 18 Exclusion criteria: Bishop score ≥ 7; cervix > 3 cm dilated; > 9 contractions per hour; estimated fetal weight < 2000 g or > 4500 g; evidence of cephalopelvic disproportion; placenta previa; unexplained vaginal bleeding; active genital herpes simplex; previous C-section or uterine surgery; chorioamnionitis; parity ≥ 6; pre-existing moderate or severe disease; contraindications to prostaglandins Age (mean ± SD): Mifepristone, 27.2 ± 5.9; placebo, 25.8 ± 5.4 Race: NR Gestational age at entry (mean ± SD): Mifepristone, 41.4 ± 0.4 weeks; placebo, 41.4 ± 0.4 weeks Dating criteria: 1) LMP confirmed by physical exam at 20 weeks or U/S no later than 26 weeks; *or* 2) U/S no later than 26 weeks Parity (mean ± SD): Mifepristone, 1.5 ± 1.4, 26% nulliparous;	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Abnormal FHR pattern 4) Birthweight 5) Admission to NICU 6) Length of stay in NICU 7) Plasma glucose, day 1 8) Plasma glucose, day 2 9) C-sections 10) Chorioamnionitis 11) Vaginal delivery in 24 hours 12) Vaginal delivery in 48 hours 13) Time to delivery 14) Time to active labor	1) Apgar score < 7 at 1 minute: Mifepristone: 15/97 (15.5%) Placebo: 7/83 (8.4%) p = 0.44 2) Apgar score < 7 at 5 minutes: Mifepristone: 2/97 (2%) Placebo: 0 p = 0.54 3) Abnormal FHR pattern: Mifepristone: 18/97 (18.6%) Placebo: 6/83 (7.2%) p = 0.34 4) Birthweight (mean ± SD) Mifepristone: 3676.57 ± 417.5 g Placebo: 3693.34 ± 501.8 p = 0.81 5) Admission to NICU: Mifepristone: 13/97 (13.4%) Placebo: 11/83 (13.3%) p = 0.98 6) Length of stay in NICU (mean ± SD): Mifepristone (n = 13): 5.5 ± 3.5 days Placebo (n = 11): 6.0 ± 4.1 days p = 0.78 7) Plasma glucose, day 1 (mean ± SD): Mifepristone: 64.8 ± 19.5 mg/dL Placebo: 66.5 ± 21.1 mg/dL p = 0.68 8) Plasma glucose, day 2 (mean ± SD): Mifepristone: 66.4 ± 19.5 mg/dL Placebo: 71.3 ± 23.1 mg/dL p =0.28 9) C-sections: Mifepristone: 9/97 (9.3%) Placebo: 18/83 (21.7%)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Sample size based on proportion of patients delivering within 48 hours. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Maternal and family medicine Length of follow-up: None	placebo, 1.1 ± 1.2, 40% nulliparous Bishop score (median, with range): Mifepristone, 2 (0 to 6); placebo, 3 (0 to 6)		p = 0.02 10) Chorioamnionitis: Mifepristone: 15/97 (15.5%) Placebo: 18/83 (21.7%) p = 0.28 11) Vaginal delivery in 24 hours: Mifepristone: 12/88 (13.6%) Placebo: 7/65 (10.8%) p = 0.60 12) Vaginal delivery in 48 hours: *Overall:* Mifepristone: 77/88 (87.5%) Placebo: 46/65 (70.8%) p = 0.01 *Among nulliparas:* Mifepristone: 15/25 (60.0%) Placebo: 10/34 (29.4%) (no p-value reported) *Among multiparas:* Mifepristone: 62/72 (86.1%) Placebo: 36/49 (73.5%) (no p-value reported) 13) Time to delivery (mean ± SD): *Overall:* Mifepristone: 2209 ± 698 minutes Placebo: 2671 ± 884 minutes p < 0.001 *Among nulliparas:* Mifepristone (n = 25): 2426 ± 804 minutes Placebo (n = 34): 3169 ± 875 minutes p = 0.002 *Among multiparas:* Mifepristone (n = 72): 2129 ± 644 minutes Placebo (n = 49): 2326 ± 714 minutes	*(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				p = 0.16 14) Time to active labor (mean ± SD): Mifepristone: 1890 ± 668 minutes Placebo: 2303 ± 806 minutes p = 0.002	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Wing, Jones, Rahall, et al., 1995	Design: RCT, randomization by table of random numbers and sealed envelopes Interventions: 1) Misoprostol (n = 68) Protocol: Misoprostol 50 µg applied intravaginally to posterior fornix. Dose repeated every 3 hours until adequate contraction pattern established (3 contractions in 10 minutes), Bishop score ≥ 8, dilation ≥ 3 cm, or SROM occurred. Maximum dose 300 µg or 6 doses 2) PGE_2 gel (n = 67) Protocol: PGE_2 gel (0.5 mg) applied intracervically every 6 hours as necessary to a maximum of 3 doses. In both groups, artifical rupture of the membranes generally performed when the cervix was 80% effaced and 3 cm dilated. If patient did not enter active labor after receiving maximum dose, had SROM without ensuing adequate contractile pattern, or had an arrest of dilatation, then IV oxytocin augmentation given (3 hours after last dose of misoprostol or ≥ 6 hours after last dose of PGE_2). Dates: Oct –Nov 1993 Location: Los Angeles, CA Setting: University hospital	No. of subjects at start: 135 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 135 Inclusion criteria: Medical or obstetric indication for induction; singleton gestation; cephalic presentation; intact membranes; Bishop score ≤ 4; reactive NST; < 4 spontaneous uterine contractions per hour Exclusion criteria: Abnormal FHR patterns; malpresentation; estimated fetal weight > 4500 g or other evidence of cephalopelvic disproportion; ruptured membranes; placenta previa or other unexplained vaginal bleeding; vasa previa; active herpes simplex infection; contraindication to prostaglandins; renal or hepatic dysfunction; suspected chorioamnionitis; previous C-section or history of uterine surgery; parity > 5 Age (mean ± SD): Misoprostol, 24.9 ± 6.9; PGE_2, 26.4 ± 6.9 Race: NR Gestational age at entry (mean ± SD): Misoprostol, 39.9 ± 2.3 weeks; PGE_2, 40.3 ± 1.9 weeks Dating criteria: NR Parity: Misoprostol / PGE_2 Nullip 52% / 48%	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Meconium aspiration syndrome 5) Admission to NICU 6) Neonatal resuscitation 7) Forceps delivery 8) Vacuum delivery 9) C-sections (overall and by indication) 10) Time to delivery 11) Vaginal delivery in 24 hours 12) Tachysystole 13) Hyperstimulation	1) Apgar score < 7 at 1 minute: Misoprostol: 9/68 (13.2%) PGE_2: 6/67 (9.0%) p = not significant 2) Apgar score < 7 at 5 minutes: Misoprostol: 1/68 (1.5%) PGE_2: 0/67 p = not significant 3) Birthweight (mean ± SD): Misoprostol: 3273.5 ± 522.4 g PGE_2: 3356.0 ± 523.0 g p = not significant 4) Meconium aspiration syndrome: Misoprostol: 3/68 (4.4%) PGE_2: 1/67 (1.5%) p < 0.05 5) Admission to NICU: Misoprostol: 13/68 (9.6%) PGE_2: 11/67 (8.1%) p = not significant 6) Neonatal resuscitation: Misoprostol: 15/68 (22.1%) PGE_2: 5/67 (7.5%) p < 0.05 7) Forceps delivery: Misoprostol: 2/68 (2.9%) PGE_2: 2/67 (3.0%) p = not significant 8) Vacuum delivery: Misoprostol: 5/68 (7.4%) PGE_2: 6/67 (8.9%) p = not significant 9) C-sections: *Overall:* Misoprostol: 10/68 (14.7%) PGE_2: 13/67 (19.4%)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (10% of total study population). Results not stratified by parity. Sample size estimates based on proportion of patients achieving “adequate labor pattern” and proportion undelivered at 24 hours. Study underpowered to detect differences in categorical variables (e.g., tachysystole). *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Type(s) of providers: NR Length of follow-up: None	Primip 16% 19% Multip 32% 33% Bishop score (median, with range): Misoprostol, 2 (0-4); PGE2, 2 (0-4) Other: Indications for induction: Oligohydramnios: 58% Preeclampsia: 14% Postterm: 10% Macrosomia: 7% Abnormal antepartum testing: 3% Rh sensitization: 2% IUGR: 1% Diabetes mellitus: 1% Chronic hypertension: 1% Other: 3%		p = not significant *For arrest disorder:* Misoprostol: 6/68 (8.8%) PGE_2: 7/67 (10.4%) (no p-value reported) *For failed induction:* Misoprostol: 3/68 (4.4%) PGE_2: 5/67 (7.5%) (no p-value reported) *For fetal distress:* Misoprostol: 1/68 (1.5%) PGE_2: 1/67 (1.5%) (no p-value reported) 10) Time to delivery (mean ± SD): *Any delivery:* Misoprostol: 1100.9 ± 751.4 minutes PGE_2: 1592.6 ± 927.5 minutes $p < 0.001$ *Vaginal delivery:* Misoprostol: 903.3 ± 482.1 minutes PGE_2: 1410.9 ± 869.1 minutes $p < 0.001$ 11) Vaginal delivery in 24 hours: Misoprostol: 48/68 (70.6%) PGE_2: 32/67 (47.8%) $p < 0.01$ 12) Tachysystole: Misoprostol: 25/68 (36.7%) PGE_2: 8/67 (11.9%) $p < 0.001$ 13) Hyperstimulation Misoprostol: 5/68 (7.4%) PGE_2: 2/67 (3.0%) p = not significant	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Wing, Ortiz-Omphroy, and Paul, 1997	Design: RCT, randomization by computer-generated random numbers and sealed envelopes Interventions: 1) PGE_2 (n = 98) Protocol: 10-mg vaginal insert place in posterior fornix. Drug released at rate of 0.3 mg per hour. Insert removed if active labor (dilation ≥ 4 cm), SROM, Bishop score ≥ 8, cervical dilation ≥ 3 cm, uterine contraction abnormality (tachysystole, hypertonus, or hyperstimulation), abnormal FHR activity, or after 24 hours. 2) Misoprostol (n = 99) Protocol: 25 µg placed in posterior vaginal fornix every 4 hours until adequate contraction pattern established (3 contractions in 10 minutes), Bishop score ≥ 8, dilation ≥ 3 cm, SROM occurred, or 24 hours passed. Maximum dose 150 µg, or 6 doses. In both groups, AROM generally performed when cervix 80% effaced and 3 cm dilated, or when dilatation > 4 cm regardless of effacement. Patients who did not enter labor after maximum dose, or had SROM without adequate labor pattern, or arrest of dilatation received oxytocin augmentation. Dates: Oct 1995 - June 1996	No. of subjects at start: 200 Dropouts: 3 (excluded from analysis due to protocol violation) Loss to follow-up: NA No. of subjects at end: 197 Inclusion criteria: Medical or obstetric indication for induction; singleton gestation; cephalic presentation; intact membranes; Bishop score ≤ 4; reactive FHR pattern; < 8 spontaneous uterine contractions per hour Exclusion criteria: Abnormal FHR pattern; malpresentation; estimated fetal weight > 4500 g or other evidence of cephalopelvic disproportion; ruptured membranes; placenta previa or other unexplained vaginal bleeding; vasa previa; active herpes simplex infection; contraindications to prostaglandins; renal or hepatic dysfunction; suspected chorioamnionitis; previous C-section or other uterine surgery; parity > 5 Age: "Similar" in two groups Race: 97% Hispanic, equally distributed between the two groups Gestational age at entry (mean ± SD): PGE_2, 39.2 ± 2.3 weeks; misoprostol, 29.5 ± 2.4 weeks Dating criteria: NR	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Neonatal resuscitation 5) Admission to NICU 6) C-sections 7) Cost of study medication (per dose) 8) Vaginal delivery within 12 and 24 hours	1) Apgar score < 7 at 1 minute: PGE_2: 11/98 (11.2%) Misoprostol: 9/99 (9.1%) p = 0.29 2) Apgar score < 7 at 5 minutes: PGE_2: 0/98 Misoprostol: 0/99 p = not significant 3) Birthweight (mean ± SD): PGE_2: 3264.6 ± 592.3 g Misoprostol: 3305.8 ± 549.3 g p = 0.61 4) Neonatal resuscitation: PGE_2: 25/98 (25.5%) Misoprostol: 29/99 (29.3%) p = 0.55 5) Admission to NICU: PGE_2: 27/98 (27.6%) Misoprostol: 30/99 (30.3%) p = 0.67 6) C-sections: PGE_2: 20/98 (20.4%) Misoprostol: 18/99 (18.2%) p = not significant 7) Cost of study medication (per dose): PGE_2: $135 per insert Misoprostol: $0.08 per 25-µg dose (no p-value reported) 8) Vaginal delivery: *Within 12 hours:* PGE_2: 19/98 (19.4%) Misoprostol: 20/99 (20.2%) p = not significant *Within 24 hours:* PGE_2: 45/98 (45.9%) Misoprostol: 51/99 (51.5%)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (13% of total study population). Results not stratified by parity. Sample size based on proportion delivering within 12 hours. Tachysystole was less frequent with misoprostol than with PGE_2 (7.1% vs. 18.4%, p = 0.02). *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Location: Los Angeles, CA Setting: University hospital Type(s) of providers: Unspecified OB/GYN; senior residents Length of follow-up: None	Parity: "Similar" in the two groups Bishop score (median, with range): 2 (0-4) in both groups Other: Indications for induction: Oligohydramnios: 43% Preeclampsia: 25% Postterm: 13% Macrosomia: 6% Diabetes mellitus: 7.5% IUGR: 3.5% Chronic hypertension: 1% Other: 1%		p = not significant	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Wing and Paul, 1996	Design: RCT, randomization by computer-generated random numbers and sealed envelopes Interventions: 1) Misoprostol, 3-hour dosing regimen (n = 261) Protocol: Misoprostol 25 µg applied in posterior vaginal fornix every 3 hours until adequate contraction pattern established (3 contractions in 10 minutes), Bishop score ≥ 8, dilation ≥ 3 cm, SROM occurred, or 24 hours passed. Maximum dose 200 µg, or 8 doses. 2) Misoprostol, 6-hour dosing regimen (n = 259) Protocol: Same as above except dosing repeated every 6 hours to a maximum of 100 µg, or 4 doses. In both groups, AROM generally performed when cervix 80% effaced and 3 cm dilated, or when dilatation > 4 cm regardless of effacement. Patients who did not enter labor after maximum dose, or had SROM without adequate labor pattern, or arrest of dilatation received oxytocin augmentation. Dates: Oct 1994 - July 1995 Location: Los Angeles, CA Setting: University hospital	No. of subjects at start: 522 Dropouts: 2 (excluded from analysis due to protocol violation) Loss to follow-up: NA No. of subjects at end: 520 Inclusion criteria: Medical or obstetric indication for induction; singleton pregnancy; cephalic presentation; intact membranes; Bishop score ≤ 4; reactive FHR pattern; < 8 spontaneous uterine contractions per hour Exclusion criteria: Abnormal FHR pattern; malpresentation; estimated fetal weight > 4500 g or other evidence of cephalopelvic disproportion; ruptured membranes; placenta previa or other unexplained vaginal bleeding; vasa previa; active herpes simplex infection; contraindications to prostaglandins; renal or hepatic dysfunction; suspected chorioamnionitis; previous C-section or other uterine surgery; parity > 5 Age: "Similar" in two groups Race: 96% Hispanic, equally distributed between the two groups Gestational age at entry (mean ± SD): 3-hour dosing, 39.6 ± 2.3 weeks; 6-hour dosing, 39.5 ± 2.3 weeks	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Neonatal resuscitation 5) Admission to NICU 6) Instrumental vaginal delivery 7) C-sections 8) Maternal adverse events 9) Tachysystole 10) Time to vaginal delivery 11) Vaginal delivery within 24 hours	1) Apgar score < 7 at 1 minute: 3-hour dosing: 31/261 (13%) 6-hour dosing: 34/259 (13%) p = not significant 2) Apgar score < 7 at 5 minutes: 3-hour dosing: 3/261 (1.5%) 6-hour dosing: 4/259 (1.5%) p = not significant 3) Birthweight (mean ± SD): 3-hour dosing: 3273 ± 565.4 g 6-hour dosing: 3267.6 ± 554.1 p = not significant 4) Neonatal resuscitation: 3-hour dosing: 90/261 (34.5%) 6-hour dosing: 83/259 (32.0%) p = not significant 5) Admission to NICU: 3-hour dosing: 61/261 (23.4%) 6-hour dosing: 54/259 (20.8%) p = not significant 6) Instrumental vaginal delivery: 3-hour dosing: 16/261 (6%) 6-hour dosing: 17/259 (6.5%) p = not significant 7) C-sections: 3-hour dosing: 53/261 (20.3%) 6-hour dosing: 55/259 (21.3%) p = not significant 8) Maternal adverse events (treatment groups not specified): One maternal death from amniotic fluid embolism, 2 cesarean hysterectomies performed for vaginal hemorrhage resulting from uterine atony.	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: - Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (13% of total study population). Results not stratified by parity. Sample size estimates based on equivalence in tachysystole. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
	Type(s) of providers: MFM, senior resident Length of follow-up: None	Dating criteria: NR Parity: "Similar" in two groups Bishop score: Median, 2 in both groups (range NR) Other: Indications for induction: Oligohydramnios: 49% Preeclampsia: 17% Postterm: 13% Macrosomia: 5% Abnormal antepartum testing: 5% Diabetes mellitus: 5% IUGR: 2% Chronic hypertension: 0.6% Rh sensitization: 0.2% Other: 3%		9) Tachysystole: 3-hour dosing: 38/261 (14.6%) 6-hour dosing: 29/259 (11.2%) p = not significant 10) Time to vaginal delivery (mean ± SD): 3-hour dosing: 903.3 ± 482.1 minutes 6-hour dosing: 1410.9 ± 869.1 minutes p < 0.05 11) Vaginal delivery within 24 hours: 3-hour dosing: 133/261 (63.9%) 6-hour dosing: 113/259 (55.4%) p = not significant	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Wing, Rahall, Jones, et al., 1995	Design: RCT, randomization by table of random numbers and sealed envelopes Interventions: 1) Misoprostol (n = 138) Protocol: Misoprostol 25-µg tablet applied intravaginally to posterior fornix. Dose repeated every 3 hours until adequate contraction pattern established or until cervical ripening or SROM occurred. Maximum dose = 200 µg, or 8 doses. 2) PGE_2 (n = 137) Protocol: PGE_2 gel (0.5 mg) applied intracervically. Dose repeated every 6 hours as necessary for a maximum of 3 doses. Dates: Feb-June 1994 Location: Los Angeles, CA Setting: University hospital Type(s) of providers: Unspecified hospital Length of follow-up: None	No. of subjects at start: 276 Dropouts: 1 (excluded from analysis due to protocol violation) Loss to follow-up: NA No. of subjects at end: 275 Inclusion criteria: Medical or obstetric indication for induction; singleton gestation; cephalic presentation; intact membranes; Bishop score ≤ 4; reactive NST; < 4 spontaneous uterine contractions per hour Exclusion criteria: Abnormal FHR patterns; malpresentation; estimated fetal weight > 4500 g or other evidence of cephalopelvic disproportion; ruptured membranes; placenta previa or other unexplained vaginal bleeding; vasa previa; active herpes simplex infection; contraindication to prostaglandins; renal or hepatic dysfunction; suspected chorioamnionitis; previous C-section or history of uterine surgery; parity > 5 Age (mean ± SD): Misoprostol, 25.8 ± 6.2; PGE_2, 26.2 ± 6.5 Race: Both groups 95% Hispanic Gestational age at entry (mean ± SD): Misoprostol, 39.7 ± 2.3 weeks; PGE_2, 40.0 ± 2.4 weeks Dating criteria: NR	1) Apgar score < 7 at 1 minute 2) Apgar score < 7 at 5 minutes 3) Birthweight 4) Admission to NICU 5) Neonatal resuscitation 6) Forceps delivery 7) Vacuum delivery 8) C-sections (overall and by indication) 9) Cost of study medication per dose 10) Time to vaginal delivery 11) Vaginal delivery within 24 hours	1) Apgar score < 7 at 1 minute: Misoprostol: 15/138 (11%) PGE_2: 9/137 (7%) p = not significant 2) Apgar score < 7 at 5 minutes: Misoprostol: 0/138 PGE_2: 0/137 p = not significant 3) Birthweight (mean ± SD): Misoprostol: 3269.7 ± 587.5 g PGE_2: 3395.0 ± 607.4 g p = not significant 4) Admission to NICU: Misoprostol: 17/138 (12%) PGE_2: 23/137 (17%) p = not significant 5) Neonatal resuscitation: Misoprostol: 44/138 (32%) PGE_2: 43/137 (31%) p = not significant 6) Forceps delivery: Misoprostol: 4/138 (3%) PGE_2: 8/137 (6%) (no p-value reported) 7) Vacuum delivery: Misoprostol: 5/138 (4%) PGE_2: 11/237 (8%) (no p-value reported) 8) C-sections: *Overall:* Misoprostol: 28/138 (20%) PGE_2: 38/137 (28%) p = not significant *For abnormal FHR:* Misoprostol: 9/138 (6.5%) PGE_2: 4/137 (3%)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: + Sample size: + Statistical tests: + Gestational age: + Dating criteria: - Bishop score: + Results not reported separately for subgroup of patients induced for postterm pregnancy (16% of total study population). Results not stratified by parity. Sample size estimate based on proportion delivering within 24 hours. *(continued on next page)*

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				(no p-value reported)	
		Parity: Miso PGE_2 Nullip 47% 47% Primip 18% 23% Multip 35% 30%		*For failed induction:* Misoprostol: 4/138 (3%) PGE_2: 27/137 (20%) (no p-value reported)	
		Bishop score: NR		*For arrest disorder:* Misoprostol: 15/138 (11%) PGE_2: 7/137 (5%) (no p-value reported)	
		Other: Indications for induction: Oligohydramnios: 40% Preeclampsia: 23% Postterm: 16% Macrosomia: 10% Diabetes mellitus: 5% Abnormal antepartum testing: 2% Chronic hypertension: 2% IUGR: 2% Other: 1%		9) Cost of study medication per dose: Misoprostol: $0.08 PGE_2: $75.00 (no p-value reported)	
				10) Time to vaginal delivery (mean ± SD): Misoprostol: 1323.0 ± 844.4 minutes PGE_2: 1532.4 ± 706.5 minutes $p < 0.05$	
				11) Vaginal delivery within 24 hours Misoprostol: 72/138 (65.5%) PGE_2: 41/137 (41.4%) $p < 0.01$	

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Wiriya-sirivaj, Vutya-vanich, and Ruangsri, 1996	Design: RCT, randomization by table of random numbers Interventions: 1) Membrane stripping (n = 61) Protocol: Membranes stripped by digital separation from lower uterine segment as far as possible. Unfavorable cervixes stretched digitally as far as possible or until stripping could be accomplished. Repeated weekly until labor or 42 completed weeks' gestation. If no labor at 42 weeks, then labor induced with prostaglandin suppository or oxytocin drip. 2) Pelvic exam (control) (n = 59) Protocol: Pelvic exam to assess Bishop score only. Repeated weekly until labor or 42 completed weeks' gestation. If no labor at 42 weeks, then labor induced with prostaglandin suppository or oxytocin drip. Dates: Oct-Nov 1994 Location: Chiang Mai, Thailand Setting: University hospital Type(s) of providers: Unspecified OB/GYN Length of follow-up: None	No. of subjects at start: 120 Dropouts: 0 Loss to follow-up: NA No. of subjects at end: 120 Inclusion criteria: Gestational age 38 weeks; vertex presentation; no size-date discrepancy; no placenta previa or low-lying placenta; ability to attend follow-up visits; intention to deliver at study hospital Exclusion criteria: Previous C-section; known medical or surgical or obstetric complication that would preclude vaginal delivery; high risk Age (mean ± SD): Stripping, 25.6 ± 4.9; control, 26.2 ± 4.9 Race: NR Gestational age at entry: 38 weeks Dating criteria: LMP; early assessment of uterine size; or U/S before 28 weeks Parity: Both groups, 56% primigravidae Bishop score (mean ± SD): Stripping, 2.3 ± 1.5; control, 2.1 ± 1.7	1) Birthweight 2) Apgar scores at 1 minute 3) Apgar scores at 5 minutes 4) Neonatal jaundice 5) Post-partum fever 6) Post-partum hemorrhage 7) Forceps-assisted delivery 8) Vacuum extraction 9) C-section 10) Proportion of patients delivering within 7 days 11) Incidence of postterm pregnancies	1) Birthweight (mean ± SD): Stripping: 3123 ± 364.8 g Control: 3078 ± 320.5 g p = not significant 2) Apgar scores at 1 minute (mean ± SD): Stripping: 9.1 ± 1.1 Control: 9.1 ± 1.2 p = not significant 3) Apgar scores at 5 minutes (mean ± SD): Stripping: 9.9 ± 0.2 Control: 9.9 ± 0.1 p = not significant 4) Neonatal jaundice: Stripping: 4/61 (6.6%) Control: 4/59 (6.8%) p = not significant 5) Post-partum fever: Stripping: 1/61 (1.6%) Control: 0 p = not significant 6) Post-partum hemorrhage: Stripping: 2/61 (3.3%) Control: 2/59 (3.4%) p = not significant 7) Forceps-assisted delivery: Stripping: 2/61 (3.3%) Control: 5/59 (8.5%) (no p-value reported) 8) Vacuum extraction: Stripping: 8/61 (13.1%) Control: 6/59 (10.2%) (no p-value reported)	QUALITY SCORE: Randomized: + Method of randomization: + Similar to likely pt pop: + Interventions described: + Mode of delivery: - Sample size: + Statistical tests: + Gestational age: + Dating criteria: + Bishop score: + Results not stratified by parity.

(continued on next page)

Evidence Table 3: Studies relevant to Key Question 3 (continued)

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
				9) C-section: Stripping: 6/61 (9.8%) Control: 3/59 (5.0%) (no p-value reported) 10) Proportion of patients delivering within 7 days: Stripping: 25/61 (41.0%) Control: 12/59 (20.3%) p = 0.014 11) Incidence of postterm pregnancies: Stripping: 1/61 (1.6%) Control: 3/59 (5.1%) p = not significant	

Bibliography

Abotalib ZM, Soltlan MH, Chowdhury N, et al. Obstetric outcome in uncomplicated prolonged pregnancy. Int J Gynecol Obstet 1996;55(3):225-30.

Acker DB, Sachs BP, Friedman EA. Risk factors for shoulder dystocia. Obstet Gynecol 1985;66(6):762-8.

ACOG. Ultrasonography in pregnancy. Technical Bulletin No. 187. 1993.

ACOG. Induction of labor. Technical Bulletin No. 217. 1995.

ACOG. Routine ultrasound in low-risk pregnancy. 1997.

ACOG. Fetal macrosomia. ACOG Practice Bulletin No. 22. 2000.

Adair CD, Weeks JW, Barrilleaux PS, et al. Labor induction with oral versus vaginal misoprostol: A randomized, double-blind trial. Am J Obstet Gynecol 1998;178:93.

Adetoro OO, Agah A. The implications of childbearing in postpubertal girls in Sokoto, Nigeria. Int J Gynaecol Obstet 1988;27(1):73-7.

Agarwal N, Mathur T, Gupta N, et al. Role of intra-amniotic instillation of betamethasone in prolonged pregnancy. Indian J Med Res 1981;73(3):335-41.

Ahlden S, Andersch B, Stigsson L, et al. Prediction of sepsis neonatorum following a full-term pregnancy. Gynecol Obstet Invest 1988;25(3):181-5.

Ahmed AI, Versi E. Prolonged pregnancy. [Review]. Curr Opin Obstet Gynecol 1993;5(5):669-74.

Al-Najashi SS, Al-Mulhim AA. Prolongation of pregnancy in multiple pregnancy. Int J Gynecol Obstet 1996;54(2):131-5.

Alberico S, Fadalti M, Grimaldi E, et al. Eligibility criteria for labor induction with prostaglandins. Clin Exp Obstet Gynecol 1997;24(2):61-6.

Alexander JM, McIntire DD, Leveno KJ. Forty weeks and beyond: pregnancy outcomes by week of gestation. Obstet Gynecol 2000;96(2):291-4.

Alexander JM, Mcintire DD, Leveno KJ. Prolonged pregnancy: induction of labor and cesarean births. Obstet Gynecol 2001;97(6911-5).

Alfirevic Z, Howarth G, Gaussmann A. Oral misoprostol for induction of labor (Cochrane Review). In: The Cochrane Library, Issue 2, 2000. Oxford: Update Software.

Alfirevic Z, Luckas M, Walkinshaw SA, et al. A randomised comparison between amniotic fluid index and maximum pool depth in the monitoring of post-term pregnancy. Br J Obstet Gynaecol 1997;104(2):207-11.

Alfirevic Z, Neilson JP. Biophysical profile for fetal assessment in high risk pregnancies (Cochrane Review). In: The Cochrane Library, Issue 2, 2001. Oxford: Update Software.

Alfirevic Z, Walkinshaw SA. Management of post-term pregnancy: to induce or not? [Review] . Br J Hosp Med 1994;52(5):218-21.

Alfirevic Z, Walkinshaw SA. A randomised controlled trial of simple compared with complex antenatal fetal monitoring after 42 weeks of gestation. Br J Obstet Gynaecol 1995;102(8):638-43.

Allott HA, Palmer CR. Sweeping the membranes: A valid procedure in stimulating the onset of labour? Br J Obstet Gynaecol 1993;100(10):898-903.

Almstrom H, Granstrom L, Ekman G. Serial antenatal monitoring compared with labor induction in post-term pregnancies. Acta Obstet Gynecol Scand 1995;74(8):599-603.

Amano K, Saito K, Shoda T, et al. Elective induction of labor at 39 weeks of gestation: A prospective randomized trial. J Obstet Gynaecol Res 1999;25(1):33-7.

Amiel-Tison C. Cerebral handicap in full-term newborns related to late pregnancy and/or labor. Eur J Obstet Gynecol Reprod Biol 1988;28(2):157-63.

Anderson T. Cochrane made simple. Routine ultrasound in early pregnancy. Pract Midwife 1998;1(12):13-4.

Anderson T. Cochrane made simple. Post-term pregnancy. Pract Midwife 1999;2(11):10-2.

Anderson T. Post-term pregnancy. [Review]. Pract Midwife 1999;2(11):10-2.

Anonymous. Task force on cerebral palsy and neonatal asphyxia. In: The Society of Obstetricians and Gynaecologists of Canada: Policy Statement, Committee Opinion, Clinical Practice Guidelines. Ottawa, ON : SOGC, 1996. 72 P.

Anonymous. Prolonged pregnancy: the management debate. BMJ 1988;297(6650):715-7.

Anonymous. Induction and augmentation of labor: ACOG Technical bulletin number 157 (replaces No. 110, November 1987). Int J Gynecol Obstet 1992;39(2):139-42.

Anonymous. The management of postpartum haemorrhage. Drug Ther Bull 1992;30(23):89-92.

Anonymous. Post-term induction: don't kill the messenger. Lancet 1992;340(8823):824.

Anonymous. Postterm pregnancy. Int J Gynecol Obstet 1996;53(1):89-90.

Anonymous. ACOG practice patterns. Management of postterm pregnancy. Number 6, October 1997. American College of Obstetricians and Gynecologists. Int J Gynaecol Obstet 1997;60(1):86-91.

Anonymous. Managing post-term pregnancy. [Review]. Drug Ther Bull 1997;35(3):17-8.

Anonymous. Electronic fetal heart rate monitoring: research guidelines for interpretation. The National Institute of Child Health and Human Development Research Planning Workshop. J Obstet Gynecol Neonat Nurs 1997;26(6):635-40.

Apgar J, Everett GA, Fitzgerald JA. Dietary zinc deprivation affects parturition and outcome of pregnancy in the ewe. Nutr Res 1993;13(3):319-30.

Arabin B, Becker R, Mohnhaupt A, et al. Prediction of fetal distress and poor outcome in prolonged pregnancy using Doppler ultrasound and fetal heart rate monitoring combined with stress tests (II). Fetal Diagn Ther 1994;9(1):1-6.

Arabin B, Snyjders R, Mohnhaupt A, et al. Evaluation of the fetal assessment score in pregnancies at risk for intrauterine hypoxia. Am J Obstet Gynecol 1993;169(3):549-54.

Arias F. Predictability of complications associated with prolongation of pregnancy. Obstet Gynecol 1987;70(1):101-6.

Ariceta JM, Matorras R, Diez J, et al. Sequential vacuum extractor and forceps delivery. J Obstet Gynaecol 1994;14(5):312-16.

Arntzen KJ, Brekke O-L, Vatten L, et al. Reduced production of PGEinf 2 and PGF(2alpha) from decidual cell cultures supplemented with N-3 polyunsaturated fatty acids. Prostaglandins Other Lipid Mediat 1998;56(2-3):183-95.

Arrowsmith S, Hamlin EC, Wall LL. Obstructed labor injury complex: Obstetric fistula formation and the multifaceted morbidity of maternal birth trauma in the developing world. Obstet Gynecol Surv 1996;51(9):568-74.

Atad J, Hallak M, Auslender R, et al. A randomized comparison of prostaglandin E2, oxytocin, and the double-balloon device in inducing labor. Obstet Gynecol 1996;87(2):223-7.

Augensen K, Bergsjø P, Eikeland T, et al. Randomised comparison of early versus late induction of labour in post-term pregnancy. BMJ 1987;294(6581):1192-5.

Ayres-de-Campos D, Bernardes J, Costa-Pereira A, et al. Inconsistencies in classification by experts of cardiotocograms and subsequent clinical decision. Br J Obstet Gynaecol 1999;106(12):1307-10.

Babinszki A, Kerenyi T, Torok O, et al. Perinatal outcome in grand and great-grand multiparity: effects of parity on obstetric risk factors. Am J Obstet Gynecol 1999;181(3):669-74.

Bailey SM, Sarmandal P, Grant JM. A comparison of three methods of assessing inter-observer variation applied to measurement of the symphysis-fundal height. Br J Obstet Gynaecol 1989;96(11):1266-71.

Baines DL, Folkesson HG, Norlin A, et al. The influence of mode of delivery, hormonal status and postnatal Oinf 2 environment on epithelial sodium channel (ENaC) expression in perinatal guinea-pig lung. J Physiol 2000;522(1):147-57.

Bakimer R, Blank M, Kosashvilli D, et al. Antiphospholipid syndrome and the idiotypic network. Lupus 1995;4(3):204-8.

Bakos O, Backstrom T. Induction of labor: a prospective, randomized study into amniotomy and oxytocin as induction methods in a total unselected population. Acta Obstet Gynecol Scand 1987;66(6):537-41.

Barrett JF, Tyrrell S, Lilford RJ. Prolonged pregnancy. BMJ 1987;294(6563):56-7.

Barss VA, Frigoletto FD, Diamond F. Stillbirth after nonstress testing. Obstet Gynecol 1985;65(4):541-4.

Bartnicki J, Ratanasiri T, Meyenburg M, et al. Postterm pregnancy: computer analysis of the antepartum fetal heart rate patterns. Int J Gynaecol Obstet 1992;37(4):243-6.

Barton DPJ, Robson MS, Turner MJ, et al. Prolonged spontaneous labour in primigravidae whose labour was actively managed: Results of an audit. J Obstet Gynaecol 1992;12(5):304-8.

Baskett TF. Gestational age and fetal biophysical assessment. Am J Obstet Gynecol 1988;158(2):332-4.

Baskett TF, Allen AC. Perinatal implications of shoulder dystocia. Obstet Gynecol 1995;86(1):14-7.

Bastian H, Keirse MJ, Lancaster PA. Perinatal death associated with planned home birth in Australia: population based study. BMJ 1998;317(7155):384-8.

Battaglia C, Artini PG, Ballestri M, et al. Hemodynamic, hematological and hemorrheological evaluation of post-term pregnancy. Acta Obstet Gynecol Scand 1995;74(5):336-40.

Battaglia C, Larocca E, Lanzani A, et al. Doppler velocimetry in prolonged pregnancy. Obstet Gynecol 1991;77(2):213-6.

Beck I, Clayton JK. Hazards of prostaglandin pessaries in postmaturity. Lancet 1982;2(8290):161.

Belizan JM, Villar J, Nardin JC, et al. Diagnosis of intrauterine growth retardation by a simple clinical method: measurement of uterine height. Am J Obstet Gynecol 1978;131(6):643-6.

Bell RJ, Permezel M, MacLennan A, et al. A randomized, double-blind, placebo-controlled trial of the safety of vaginal recombinant human relaxin for cervical ripening. Obstet Gynecol 1993;82(3):328-33.

Benacerraf BR, Frigoletto FD. Fetal respiratory movements: only part of the biophysical profile. Obstet Gynecol 1986;67(4):556-7.

Benedetti TJ, Easterling T. Antepartum testing in postterm pregnancy. J Reprod Med 1988;33(3):252-8.

Benifla JL, Goffinet F, Darai E, et al. Emergency cervical cerclage after 20 weeks' gestation: A retrospective study of 6 years' practice in 34 cases. Fetal Diagn Ther 1997;12(5):274-8.

Bennett BB. Uterine rupture during induction of labor at term with intravaginal misoprostol. Obstet Gynecol 1997;89(5 Pt 2):832-3.

Berghella V, Rogers RA, Lescale K. Stripping of membranes as a safe method to reduce prolonged pregnancies. Obstet Gynecol 1996;87(6):927-31.

Bergsjo P, Bakketeig LS, Eikhom SN. Case-control analysis of post-term induction of labor. Acta Obstet Gynecol Scand 1982;61(4):317-24.

Bergsjø P, Huang GD, Yu SQ, et al. Comparison of induced versus non-induced labor in post-term pregnancy. A randomized prospective study. Acta Obstet Gynecol Scand 1989;68(8):683-7.

Bernardes J, Costa-Pereira A, Ayres-de-Campos D, et al. Evaluation of interobserver agreement of cardiotocograms. Int J Gynaecol Obstet 1997;57(1):33-7.

Bernstein IM, Watson M, Simmons GM, et al. Maternal brain death and prolonged fetal survival. Obstet Gynecol 1989;74(3 II SUPPL.):434-7.

Berzosa J, Martinez-Guisasola J, Guerrero M, et al. Application of prostaglandin Einf 2 gel in post-term pregnancy (> 287 days). Clin Invest Ginecol Obstet 1996;23(3):101-5.

Bishop EH. Pelvic scoring for elective induction. Obstet Gynecol 1964;24:266-8.

Blakemore KJ, Qin NG, Petrie RH, et al. A prospective comparison of hourly and quarter-hourly oxytocin dose increase intervals for the induction of labor at term. Obstet Gynecol 1990;75(5):757-61.

Blanchette HA, Nayak S, Erasmus S. Comparison of the safety and efficacy of intravaginal misoprostol (prostaglandin E1) with those of dinoprostone

(prostaglandin E2) for cervical ripening and induction of labor in a community hospital. Am J Obstet Gynecol 1999;180(6 Pt 1):1551-9.

Blanco JD, Collins M, Willis D, et al. Prostaglandin E2 gel induction of patients with a prior low transverse cesarean section. Am J Perinatol 1992;9(2):80-3.

Bobby PD, Divon MY. Fetal testing in postdates. Curr Opin Obstet Gynecol 1997;9(2):79-82.

Bobrowski RA, Bottoms SF. Underappreciated risks of the elderly multipara. Am J Obstet Gynecol 1995;172(6):1764-7.

Bochner CJ, Medearis AL, Davis J, et al. Antepartum predictors of fetal distress in postterm pregnancy. Am J Obstet Gynecol 1987;157(2):353-8.

Bochner CJ, Medearis AL, Ross MG, et al. The role of antepartum testing in the management of postterm pregnancies with heavy meconium in early labor. Obstet Gynecol 1987;69(6):903-7.

Bochner CJ, Williams J 3d, Castro L, et al. The efficacy of starting postterm antenatal testing at 41 weeks as compared with 42 weeks of gestational age. Am J Obstet Gynecol 1988;159(3):550-4.

Bors-Koefoed R, Zylstra S, Resseguie LJ, et al. A cost analysis of ambulatory antenatal testing in a tertiary care center. J Matern Fetal Invest 1993;2(1):5-9.

Boulvain M, Fraser WD, Marcoux S, et al. Does sweeping of the membranes reduce the need for formal induction of labour? A randomised controlled trial. Br J Obstet Gynaecol 1998;105(1):34-40.

Boulvain M, Irion O. Stripping/sweeping the membranes for inducing labour or preventing post-term pregnancy (Cochrane Review). In: The Cochrane Library, Issue 2, 2001. Oxford: Update Software.

Boulvain M, Irion O, Lohse C, et al. Mechanical methods for inducing labour (Cochrane Protocol). In: The Cochrane Library, Issue 4, 2000. Oxford: Update Software. (withdrawn).

Boulvain M, Irion O, Marcoux S, et al. Sweeping of the membranes to prevent post-term pregnancy and to induce labour: a systematic review. [Review]. Br J Obstet Gynaecol 1999;106(5):481-5.

Boulvain M, Stan C, Irion O. Membrane sweeping for induction of labour (Cochrane Review). In: The Cochrane Library, Issue 2, 2001. Oxford: Update Software.

Bourgeois FJ, Harbert GM, Andersen WA, et al. Early versus late tocolytic treatment of preterm premature membrane rupture. Am J Obstet Gynecol 1988;159(3):742-8.

Boyd ME, Usher RH, McLean FH. Fetal macrosomia: prediction, risks, proposed management. Obstet Gynecol 1983;61(6):715-22.

Boyd ME, Usher RH, McLean FH, et al. Obstetric consequences of postmaturity. Am J Obstet Gynecol 1988;158(2):334-8.

Boylan P, McParland P. Fetal assessment in postterm pregnancy. [Review]. Curr Opin Obstet Gynecol 1991;3(1):41-4.

Brace RA, Wolf EJ. Normal amniotic fluid volume changes throughout pregnancy. Am J Obstet Gynecol 1989;161(2):382-8.

Bracero LA, Roshanfekr D, Byrne DW. Analysis of antepartum fetal heart rate tracing by physician and computer. J Matern Fetal Med 2000;9(3):181-5.

Brar HS, Horenstein J, Medearis AL, et al. Cerebral, umbilical, and uterine resistance using Doppler velocimetry in postterm pregnancy. J Ultrasound Med 1989;8(4):187-91.

Brennand JE, Calder AA, Leitch CR, et al. Recombinant human relaxin as a cervical ripening agent. Br J Obstet Gynaecol 1997;104(7):775-80.

Bresadola M, Lo Mastro F, Arena V, et al. Prognostic value of biophysical profile score in post-date pregnancy. Clin Exp Obstet Gynecol 1995;22(4):330-8.

Bricker L, Luckas M. Amniotomy alone for induction of labor (Cochrane Review). In: The Cochrane Library, Issue 1, 2001. Oxford: Update Software.

Bricker L, Neilson JP. Routine Doppler ultrasound in pregnancy (Cochrane Review). In: The Cochrane Library, Issue 3, 2000. Oxford: Update Software.

Bricker L, Neilson JP. Routine ultrasound in late pregnancy (after 24 weeks gestation) (Cochrane Review). In: The Cochrane Library, Issue 3, 2000.

Oxford: Update Software.

Brooks GG, Lewis DF, Gallaspy JW, et al. Labor in the gravida with 10 or more years between pregnancies. J Reprod Med Obstet Gynecol 1992;37(4):336-8.

Brost BC, Newman RB, Hendricks SK, et al. Effect of hemodialysis on serum progesterone level in pregnant women. Am J Kidney Dis 1999;33(5):917-9.

Brown VA, Sawers RS, Parsons RJ, et al. The value of antenatal cardiotocography in the management of high-risk pregnancy: a randomized controlled trial. Br J Obstet Gynaecol 1982;89(9):716-22.

Buser D, Mora G, Arias F. A randomized comparison between misoprostol and dinoprostone for cervical ripening and labor induction in patients with unfavorable cervices. Obstet Gynecol 1997;89(4):581-5.

Buttino LT, Garite TJ. Intracervical prostaglandin in postdate pregnancy. A randomized trial. J Reprod Med 1990;35(2):155-8.

Cabbad MF, Minkoff H, Faustin D. Fetal heart rate decelerations after oxytocin infusion in an abdominal pregnancy. Obstet Gynecol 1985;66(3 Suppl):2S-4S.

Cammu H, Haitsma V. Sweeping of the membranes at 39 weeks in nulliparous women: a randomised controlled trial. Br J Obstet Gynaecol 1998;105(1):41-4.

Campbell MK. Factors affecting outcome in post-term birth. Curr Opin Obstet Gynecol 1997;9(6):356-60.

Campbell MK, Ostbye T, Irgens LM. Post-term birth: Risk factors and outcomes in a 10-year cohort of Norwegian births. Obstet Gynecol 1997;89(4):543-8.

Cardozo L. Is routine induction of labour at term ever justified? BMJ 1993;306(6881):840-1.

Cardozo L, Fysh J, Pearce JM. Prolonged pregnancy: the management debate. BMJ 1986;293(6554):1059-63.

Cario GM. Conservative management of prolonged pregnancy using fetal heart rate monitoring only: a prospective study. Br J Obstet Gynaecol 1984;91(1):23-30.

Carlan SJ, Angel JL, Knuppel RA. Shoulder dystocia. Am Fam Physician 1991;43(4):1307-11.

Carlomagno G, Candussi G, Zavino S, et al. Postmaturity: How far is it a clinical entity in its own right? Clin Exp Obstet Gynecol 1996;23(1):41-7.

Carr MH, Towers CV, Eastenson AR, et al. Prolonged bedrest during pregnancy: Does the risk of deep vein thrombosis warrant the use of routine heparin prophylaxis? J Matern Fetal Med 1997;6(5):264-7.

Castro LC, Arora CP, Roll KE, et al. Perinatal factors influencing atrial natriuretic peptide levels in umbilical arterial plasma at the time of delivery [published erratum appears in Am J Obstet Gynecol 1989 Nov;161(5):1238]. Am J Obstet Gynecol 1989;161(3):623-7.

Center for Reviews and Dissemination Reviewers. Antenatal ultrasound scanning. Database of Abstracts of Reviews of Effectiveness Issue 1, 2000 .

Center for Reviews and Dissemination Reviewers. Misoprostol for cervical ripening and labor induction: a meta-analysis. Database of Abstracts of Reviews of Effectiveness Issue 1, 2000 .

Center for Reviews and Dissemination Reviewers. Umbilical Doppler velocimetry: results of controlled trials in high risk and low risk populations. Database of Abstracts of Reviews of Effectiveness Issue 1, 2000 .

Chang CH, Chang FM. Randomized comparison of misoprostol and dinoprostone for preinduction cervical ripening and labor induction. J Formos Med Assoc 1997;96(5):366-9.

Chang TC, Tan KT, Neow P, et al. Computerised analysis of foetal heart rate variation: Prediction of adverse perinatal outcome in patients undergoing prostaglandin induction of labour at term. Ann Acad Med Singapore 1997;26(6):772-5.

Chari RS, Friedman SA, O'Brien JM, et al. Daily antenatal testing in women with severe preeclampsia. Am J Obstet Gynecol 1995;173(4):1207-10.

Chatterjee MS, Ramchandran K, Ferlita J, et al. Prostaglandin E2 (PGE2) vaginal gel for cervical ripening. Eur J Obstet Gynecol Reprod Biol 1991;38(3):197-202.

Chauhan SP, Hendrix NW, Magann EF, et al. Limitations of clinical and sonographic estimates of birth weight: experience with 1034 parturients. Obstet Gynecol 1998;91(1):72-7.

Chauhan SP, Lutton PM, Bailey KJ, et al. Intrapartum clinical, sonographic, and parous patients' estimates of newborn birth weight. Obstet Gynecol 1992;79(6):956-8.

Chauhan SP, Magann EF, Sullivan CA, et al. Amniotic fluid index as an admission test may increase the incidence of cesarean section in a community hospital. J Matern Fetal Invest 1994;4(4):233-6.

Chauhan SP, Sullivan CA, Lutton TC, et al. Parous patients' estimate of birth weight in postterm pregnancy. J Perinatol 1995;15(3):192-4.

Chauhan SP, Sullivan CA, Magann EF, et al. Estimate of birthweight among post-term pregnancy: Clinical versus sonographic. J Matern Fetal Med 1994;3(5):208-11.

Chayen B, Tejani N, Verma U. Induction of labor with an electric breast pump. J Reprod Med 1986;31(2):116-8.

Chervenak JL. Macrosomia in the postdates pregnancy. [Review] . Clin Obstet Gynecol 1992;35(1):151-5.

Chervenak JL, Divon MY, Hirsch J, et al. Macrosomia in the postdate pregnancy: is routine ultrasonographic screening indicated? Am J Obstet Gynecol 1989;161(3):753-6.

Chez RA, Jonas WB. Complementary and alternative medicine. Part I: Clinical studies in obstetrics. Obstet Gynecol Surv 1997;52(11):704-8.

Chi IC, Agoestina T, Harbin J. Maternal mortality at twelve teaching hospitals in Indonesia-an epidemiologic analysis. Int J Gynaecol Obstet 1981;19(4):259-66.

Choe JK, Baggish MS. Hysteroscopic treatment of septate uterus with Neodymium-YAG laser. Fertil Steril 1992;57(1):81-4.

Christensen KK, Ingemarsson I, Leideman T, et al. Effect of ritodrine on labor after premature rupture of the membranes. Obstet Gynecol 1980;55(2):187-90.

Chuck FJ, Huffaker BJ. Labor induction with intravaginal misoprostol versus intracervical prostaglandin E2 gel (Prepidil gel): randomized comparison. Am J Obstet Gynecol 1995;173(4):1137-42.

Cibils LA. On prolonged pregnancy. Am J Obstet Gynecol 1995;172(4 Pt 1):1321-2.

Cibils LA, Votta R. Clinical significance of fetal heart rate patterns during labor. IX: Prolonged pregnancy. J Perinat Med 1993;21(2):107-16.

Clark A, Cook V, Hill P, et al. Cervical ripening and labor induction: misoprostol vs dinoprostone. Am J Obstet Gynecol 1998;178:S30.

Clark MJ. Use of the oxytocin challenge test in the management of postdate pregnancy. J Am Osteopath Assoc 1980;79(10):632-5.

Clark SL. Intrapartum management of the postdate patient. [Review]. Clin Obstet Gynecol 1989;32(2):278-84.

Clarkson CP, Magann EF, Siddique SA, et al. Hematological complications of Gaucher's disease in pregnancy. Mil Med 1998;163(7):499-501.

Clausson B, Cnattingius S, Axelsson O. Outcomes of post-term births: the role of fetal growth restriction and malformations. Obstet Gynecol 1999;94(5 Pt 1):758-62.

Clement D, Schifrin BS, Kates RB. Acute oligohydramnios in postdate pregnancy. Am J Obstet Gynecol 1987;157(4 Pt 1):884-6.

Clifford SH. Postmaturity--with placental dysfunction. J Pediatr 1954;44(1):1-13.

Cnattingius S, Taube A. Stillbirths and rate of neonatal deaths in 76,761 postterm pregnancies in Sweden, 1982-1991; a register study. Acta Obstet Gynecol Scand 1998;77(5):582-3.

Collins E. Maternal and fetal effects of acetaminophen and salicylates in pregnancy. [Review]. Obstet Gynecol 1981;58(5 Suppl):57S-62S.

Combs CA, Singh NB, Khoury JC. Elective induction versus spontaneous labor after sonographic diagnosis of fetal macrosomia. Obstet Gynecol 1993;81(4):492-6.

Cosner KR, Dougherty MC, Bishop KR. Dynamic characteristics of the circumvaginal muscles during pregnancy and the postpartum. J Nurse Midwifery 1991;36(4):221-5.

Cotzias CS, Paterson-Brown S, Fisk NM. Prospective risk of unexplained stillbirth in singleton pregnancies at term: population based analysis. BMJ 1999;319(7205):287-8.

Cousins ME. The reproductive risk potential of prolonged gestation. [Review]. Issues Health Care Women 1981;3(3):139-50.

Crane J, Bennett K, Young D, et al. The effectiveness of sweeping membranes at term: a randomized trial. Obstet Gynecol 1997;89(4):586-90.

Crowley P. Interventions for preventing or improving the outcome of delivery at or beyond term (Cochrane Review). In: The Cochrane Library, Issue 2, 2000. Oxford: Update Software.

Crowley P, O'Herlihy C, Boylan P. The value of ultrasound measurement of amniotic fluid volume in the management of prolonged pregnancies. Br J Obstet Gynaecol 1984;91(5):444-8.

Crowther CA. Prevention of preterm birth in multiple pregnancy. Bailliere's Clin Obstet Gynaecol 1998;12(1):67-75.

Cruikshank DP. Amniocentesis for determination of fetal maturity. [Review]. Clin Obstet Gynecol 1982;25(4):773-85.

Cruikshank DP, Linyear AS. Term stillbirth: causes and potential for prevention in Virginia. Obstet Gynecol 1987;69(6):841-4.

Crump WJ. Postdate pregnancy in a network of community hospitals: management and outcome. J Fam Pract 1988;26(1):41-4.

Crump WJ, Smith CW. The postdate pregnancy. When to wait, when to induce labor. Postgrad Med 1986;80(5):291-3, 296-7.

Csapo AI, Eskola J, Ruttner Z. The biological meaning of progesterone levels. Prostaglandins 1980;19(2):203-11.

Cucco C, Osborne MA, Cibils LA. Maternal-fetal outcomes in prolonged pregnancy. Am J Obstet Gynecol 1989;161(4):916-20.

Cundy T, Gamble G, Townend K, et al. Perinatal mortality in Type 2 diabetes mellitus. Diabet Med 2000;17(1):33-9.

Curtis PD, Matthews TG, Clarke TA, et al. Neonatal seizures: the Dublin Collaborative Study. Arch Dis Child 1988;63(9):1065-8.

Davis DL, Stewart M, Harmon RJ. Postponing pregnancy after perinatal death: perspectives on doctor advice. J Am Acad Child Adolesc Psychiatry 1989;28(4):481-7.

Davis L. The use of castor oil to stimulate labor in patients with premature rupture of membranes. J Nurse Midwifery 1984;29(6):366-70.

Deayton JM, Young IR, Hollingworth SA, et al. Effect of late hypothalamo-pituitary disconnection on the development of the HPA axis in the ovine fetus and the initiation of parturition. J Neuroendocrinol 1994;6(1):25-31.

Delpapa EH, Mueller-Heubach E. Pregnancy outcome following ultrasound diagnosis of macrosomia. Obstet Gynecol 1991;78(3 Pt 1):340-3.

Dennis KJ. Prolonged pregnancy: the management debate. BMJ 1986;293(6559):1434-5.

Devoe L, Golde S, Kilman Y, et al. A comparison of visual analyses of intrapartum fetal heart rate tracings according to the new national institute of child health and human development guidelines with computer analyses by an automated fetal heart rate monitoring system. Am J Obstet Gynecol 2000;183(2):361-6.

Devoe LD, Sholl JS. Postdates pregnancy. Assessment of fetal risk and obstetric management. J Reprod Med 1983;28(9):576-80.

Divon MY, Haglund B, Nisell H, et al. Fetal and neonatal mortality in the postterm pregnancy: the impact of gestational age and fetal growth restriction. Am J Obstet Gynecol 1998;178(4):726-31.

Dizon-Townson D, Ward K. The genetics of labor. [Review]. Clin Obstet Gynecol 1997;40(3):479-84.

Doany W, McCarty J. Outpatient management of the uncomplicated postdate pregnancy with intravaginal prostaglandin E2 gel and membrane stripping. J Matern Fetal Med 1997;6(2):71-8.

Donker DK, van Geijn HP, Hasman A. Interobserver

variation in the assessment of fetal heart rate recordings. Eur J Obstet Gynecol Reprod Biol 1993;52(1):21-8.

Dove D, Johnson P. Oral evening primrose oil: its effect on length of pregnancy and selected intrapartum outcomes in low-risk nulliparous women. J Nurse Midwifery 1999;44(3):320-4.

Druzin ML, Karver ML, Wagner W, et al. Prospective evaluation of the contraction stress and nonstress tests in the management of post-term pregnancy. Surg Gynecol Obstet 1992;174(6):507-12.

Duff C, Sinclair M. Exploring the risks associated with induction of labour: a retrospective study using the NIMATS database. Northern Ireland Maternity System. J Adv Nurs 2000;31(2):410-7.

Dunn PA, Rogers D, Halford K. Transcutaneous electrical nerve stimulation at acupuncture points in the induction of uterine contractions. Obstet Gynecol 1989;73(2):286-90.

Dyson DC. Fetal surveillance vs. labor induction at 42 weeks in postterm gestation. J Reprod Med 1988;33(3):262-70.

Dyson DC, Miller PD, Armstrong MA. Management of prolonged pregnancy: induction of labor versus antepartum fetal testing. Am J Obstet Gynecol 1987;156(4):928-34.

Ecker JL, Greenberg JA, Norwitz ER, et al. Birth weight as a predictor of brachial plexus injury. Obstet Gynecol 1997;89(5 Pt 1):643-7.

Eden RD. Postdate pregnancy: antenatal assessment of fetal well-being. [Review]. Clin Obstet Gynecol 1989;32(2):235-44.

Eden RD, Gergely RZ, Schifrin BS, et al. Comparison of antepartum testing schemes for the management of the postdate pregnancy. Am J Obstet Gynecol 1982;144(6):683-92.

Eden RD, Seifert LS, Kodack LD, et al. A modified biophysical profile for antenatal fetal surveillance. Obstet Gynecol 1988;71(3 Pt 1):365-9.

Eden RD, Seifert LS, Winegar A, et al. Perinatal characteristics of uncomplicated postdate pregnancies. Obstet Gynecol 1987;69(3 Pt 1):296-9.

Eden RD, Seifert LS, Winegar A, et al. Postdate pregnancies: a review of 46 perinatal deaths. Am J Perinatol 1987;4(4):284-7.

Edozien LC. What do maternity statistics tell us about induction of labour? J Obstet Gynaecol 1999;19(4):343-44.

Eganhouse DJ. Monitoring the postdate pregnancy, part I: antepartum issues. Mother Baby J 2000;5(1):7-10, 26-9.

Eganhouse DJ. Monitoring the postdate pregnancy, part II: intrapartum issues. Mother Baby J 2000;5(2):5-20.41-4.

Egarter C, Kofler E, Fitz R, et al. Is induction of labor indicated in prolonged pregnancy? Results of a prospective randomised trial. Gynecol Obstet Invest 1989;27(1):6-9.

Egarter CH, Husslein PW, Rayburn WF. Uterine hyperstimulation after low-dose prostaglandin E2 therapy: tocolytic treatment in 181 cases. Am J Obstet Gynecol 1990;163(3):794-6.

Ekman G, Perssen PH, Ulmsten U, et al. The impact on labor induction of intracervically applied PGEinf 2-gel, related to gestational age in patients with an unripe cervix. Acta Obstet Gynecol Scand 1983;62(Suppl. 113):173-5.

Ekman G, Persson PH, Ulmsten U. Induction of labor in postterm pregnant women. Int J Gynaecol Obstet 1986;24(1):47-52.

El-Azeem S, Samuels P, Welch G, et al. Term labor induction with PGE1 isoprostol versus PGE2 dinoprostone. Am J Obstet Gynecol 1997;176:S113.

el-Damarawy H, el-Sibaie F, Tawfik TA. Antepartum fetal surveillance in post-date pregnancy. Int J Gynaecol Obstet 1993;43(2):145-50.

El-Sherbiny MT, El-Gharieb IH, Gewely HA. Vaginal misoprostol for induction of labor: 25 vs. 50 µg dose regimen. Int J Gynecol Obstet 2001;72:25-30.

El-Torkey M, Grant JM. Sweeping of the membranes is an effective method of induction of labour in prolonged pregnancy: a report of a randomized trial. Br J Obstet Gynaecol 1992;99(6):455-8.

Elliott CL, Brennand JE, Calder AA. The effects of

mifepristone on cervical ripening and labor induction in primigravidae. Obstet Gynecol 1998;92(5):804-9.

Elliott JP, Flaherty JF. The use of breast stimulation to prevent postdate pregnancy. Am J Obstet Gynecol 1984;149(6):628-32.

Elpek G, Uner M, Trak B, et al. Intravaginal naproxen in preterm labour: A preliminary study. J Obstet Gynaecol 1998;18(3):236-7.

Engstrom JL, Sittler CP, Swift KE. Fundal height measurement. Part 5—The effect of clinician bias on fundal height measurements. J Nurse Midwifery 1994;39(3):130-41.

Escudero F, Contreras H. A comparative trial of labor induction with misoprostol versus oxytocin. Int J Gynaecol Obstet 1997;57(2):139-43.

Essel JK, Opai-Tetteh ET. Macrosomia - Maternal and fetal risk factors. S Afr Med J 1995;85(1):43-6.

Evans MI, Dougan MB, Moawad AH, et al. Ripening of the human cervix with porcine ovarian relaxin. Am J Obstet Gynecol 1983;147(4):410-4.

Ewigman BG, Crane JP, Frigoletto FD, et al. Effect of prenatal ultrasound screening on perinatal outcome. N Engl J Med 1993;329(12):821-7.

Fabre E, Gonzalez de Aguero R, de Agustin JL, et al. Perinatal mortality in term and post-term births. J Perinat Med 1996;24(2):163-9.

Facchinetti F, Neri I, Genazzani AR. Factors predicting labour onset in patients treated with prostaglandin Einf 2 for cervical ripening. Eur J Obstet Gynecol Reprod Biol 1995;60(2):129-32.

Fairlie FM, Lang GD, Greer IA, et al. Umbilical artery Doppler flow velocity waveforms and maternal prostaglandin E2 and F2 alpha metabolite concentrations during cervical ripening with prostaglandin E2. Eur J Obstet Gynecol Reprod Biol 1990;37(1):7-13.

Fait G, Grisaru D, Shenhav M, et al. Balloon catheter with extra-amniotic saline instillation: a method of induction in pregnancies at 41 or more gestational weeks. Aust N Z J Obstet Gynaecol 1997;37(2):174-6.

Farah LA, Sanchez-Ramos L, Rosa C, et al. Randomized trial of two doses of the prostaglandin E1 analog misoprostol for labor induction. Am J Obstet Gynecol 1997;177(2):364-9.

Farmakides G, Schulman H, Ducey J, et al. Uterine and umbilical artery Doppler velocimetry in postterm pregnancy. J Reprod Med 1988;33(3):259-61.

Farmakides G, Schulman H, Winter D, et al. Prenatal surveillance using nonstress testing and Doppler velocimetry. Obstet Gynecol 1988;71(2):184-7.

Farooqi A, Holmgren PA, Engberg S, et al. Survival and 2-year outcome with expectant management of second- trimester rupture of membranes. Obstet Gynecol 1998;92(6):895-901.

Feldman GB. Prospective risk of stillbirth. Obstet Gynecol 1992;79(4):547-53.

Flamm BL. Vaginal birth after cesarean section: controversies old and new. Clin Obstet Gynecol 1985;28(4):735-44.

Flamm BL. Was the death of this baby avoidable? Birth 1990;17(4):226-7.

Fleischer A, Schulman H, Farmakides G, et al. Antepartum nonstress test and the postmature pregnancy. Obstet Gynecol 1985;66(1):80-3.

Fletcher H, Mitchell S, Frederick J, et al. Intravaginal misoprostol versus dinoprostone as cervical ripening and labor-inducing agents. Obstet Gynecol 1994;83(2):244-7.

Fletcher HM, Mitchell S, Simeon D, et al. Intravaginal misoprostol as a cervical ripening agent. Br J Obstet Gynaecol 1993;100(7):641-4.

Flynn AM, Kelly J, Mansfield H, et al. A randomized controlled trial of non-stress antepartum cardiotocography. Br J Obstet Gynaecol 1982;89(6):427-33.

Forouzan I, Cohen AW. Can umbilical and arcuate artery Doppler velocimetry predict fetal distress among prolonged pregnancies? J Ultrasound Med 1991;10(1):15-7.

Francome C, Savage W. Caesarean section in Britain and the United States 12% or 24%: is either the right rate? Soc Sci Med 1993;37(10):1199-218.

Franz WB 3d. Fetal assessment. Prim Care 1983;10(2):173-203.

Freeman RK, Garite TJ, Mondanlou H, et al. Postdate pregnancy: utilization of contraction stress testing for primary fetal surveillance. Am J Obstet Gynecol 1981;140(2):128-35.

Friedrichs PE. An active management approach to the postdate fetus with a reactive nonstress test and fetal heart rate decelerations. Obstet Gynecol 1988;71(4):657-8.

Friesen CD, Miller AM, Rayburn WF. Influence of spontaneous or induced labor on delivering the macrosomic fetus. Am J Perinatol 1995;12(1):63-6.

Froen JF, Arnestad M, Frey K, et al. Risk factors for sudden intrauterine unexplained death: epidemiologic characteristics of singleton cases in Oslo, Norway, 1986-1995. American Journal of Obstetrics & Gynecology 2001;184(4):694-702.

Frydman R, Lelaidier C, Baton-Saint-Mleux C, et al. Labor induction in women at term with mifepristone (RU 486): a double-blind, randomized, placebo-controlled study. Obstet Gynecol 1992;80(6):972-5.

Gabbe SG. Effects of identifying a high risk population. Diabetes Care 1980;3(3):486-8.

Gaffney G, Flavell V, Johnson A, et al. Cerebral palsy and neonatal encephalopathy. Arch Dis Child Fetal & Neonatal Edition 1994;70(3):F195-200.

Gardosi J, Vanner T, Francis A. Gestational age and induction of labour for prolonged pregnancy. Br J Obstet Gynaecol 1997;104(7):792-7.

Garry D, Figueroa R, Guillaume J, et al. Use of castor oil in pregnancies at term. Altern Ther Health Med 2000;6(1):77-9.

Gauthier RJ, Griego BD, Goebelsmann U. Estriol in pregnancy. VII. Unconjugated plasma estriol in prolonged gestation. Am J Obstet Gynecol 1981;139(4):382-9.

Gebhart F. Induced labor, along with use of labor drugs, on the rise. Drug Topics 2000;144(12):54.

Gegor CL, Paine LL, Johnson TR. Antepartum fetal assessment. A nurse-midwifery perspective. [Review]. J Nurse Midwifery 1991;36(3):153-67.

Gerhard I, Postneek F. Auricular acupuncture in the treatment of female infertility. Gynecol Endocrinol 1992;6(3):171-81.

Ghidini A, Doria V, Kirn V, et al. Successful outcome after antibiotic treatment of postamniocentesis membrane rupture and chorioamnionitis in multiple pregnancy. Am J Perinatol 1999;16(8):403-6.

Giacalone P-L, Daures J-P, Faure J-M, et al. The effects of mifepristone on uterine sensitivity to oxytocin and on fetal heart rate patterns. Eur J Obstet Gynecol Reprod Biol 2001;97:30-4.

Giacalone PL, Targosz V, Laffargue F, et al. Cervical ripening with mifepristone before labor induction: A randomized study. Obstet Gynecol 1998;92(4 I):487-92.

Gibb DM, Cardozo LD, Studd JW, et al. Prolonged pregnancy: is induction of labour indicated? A prospective study. Br J Obstet Gynaecol 1982;89(4):292-5.

Gibbs RS, Castillo MS, Rodgers PJ. Management of acute chorioamnionitis. Am J Obstet Gynecol 1980;136(6):709-13.

Gibbs RS, Eschenbach DA. Use of antibiotics to prevent preterm birth. Am J Obstet Gynecol 1997;177(2):375-80.

Gichangi PB, Ndinya-Achola JO, Ombete J, et al. Antimicrobial prophylaxis in pregnancy: A randomized, placebo-controlled trial with cefetamet-pivoxil in pregnant women with a poor obstetric history. Am J Obstet Gynecol 1997;177(3):680-4.

Gilbert L, Porter W, Brown VA. Postpartum haemorrhage—a continuing problem. Br J Obstet Gynaecol 1987;94(1):67-71.

Gilbert WM, Nesbitt TS, Danielsen B. Associated factors in 1611 cases of brachial plexus injury. Obstet Gynecol 1999;93(4):536-40.

Gilby JR, Williams MC, Spellacy WN. Fetal abdominal circumference measurements of 35 and 38 cm as predictors of macrosomia. J Reprod Med 2000;45:936-8.

Giles WB. Vascular Doppler techniques. Obstet Gynecol Clin North Am 1999;26(4):595-606.

Gilson GJ, O'Brien ME, Vera RW, et al. Prolonged pregnancy and the biophysical profile. A birthing center perspective. J Nurse Midwifery 1988;33(4):171-7.

Gimovsky ML, Bruce SL. Aspects of FHR tracings as warning signals. Clin Obstet Gynecol 1986;29(1):51-63.

Gjessing HK, Skjaerven R, Wilcox AJ. Errors in gestational age: evidence of bleeding early in pregnancy. Am J Public Health 1999;89(2):213-8.

Goer H. Postterm pregnancy: putting the merits of a policy of induction of labor into perspective. Birth 1996;23(3):180-1.

Goeree R, Hannah M, Hewson S. Cost-effectiveness of induction of labour versus serial antenatal monitoring in the Canadian Multicentre Postterm Pregnancy Trial. CMAJ 1995;152(9):1445-50.

Gohar J, Mazor M, Leiberman JR. GnRH in pregnancy. Arch Gynecol Obstet 1996;259(1):1-6.

Goharkhay N, Stanczyk FZ, Gentzschein E, et al. Plasma prostaglandin E2 metabolite levels during labor induction with a sustained-release prostaglandin E2 vaginal insert. J Soc Gynecol Investig 2000;7:338-42.

Goldberg J, Newman RB, Rust PF. Interobserver reliability of digital and endovaginal ultrasonographic cervical length measurements. Am J Obstet Gynecol 1997;177(4):853-8.

Gonen O, Rosen DJ, Dolfin Z, et al. Induction of labor versus expectant management in macrosomia: a randomized study. Obstet Gynecol 1997;89(6):913-7.

Gonen R, Spiegel D, Abend M. Is macrosomia predictable, and are shoulder dystocia and birth trauma preventable? Obstet Gynecol 1996;88(4 Pt 1):526-9.

Goni S, Sawhney H, Gopalan S. Oxytocin induction of labor: a comparison of 20- and 60-min dose increment levels. Int J Gynaecol Obstet 1995;48(1):31-6.

Goodlin RC. Prolongation of pregnancy and survival of second twin. Am J Obstet Gynecol 1990;163(1 Pt 1):270-1.

Gottschall D, Borgida AF, Feldman DM, et al. Preinduction cervical ripening comparing 50 and 100 mcg of misoprostol. Am J Obstet Gynecol 1998;178:S93.

Gottschall DS, Borgida AF, Mihalek JJ, et al. A randomized clinical trial comparing misoprostol with prostaglandin E2 gel for preinduction cervical ripening. Am J Obstet Gynecol 1997;177(5):1067-70.

Granados JL. Survey of the management of postterm pregnancy. Obstet Gynecol 1984;63(5):651-3.

Grant A, Hepburn M. Merits of an individualized approach to fetal movement counting compared with fixed-time and fixed-number methods. Br J Obstet Gynaecol 1984;91(11):1087-90.

Grant JM. The fetal heart rate trace is normal, isn't it? Observer agreement of categorical assessments. Lancet 1991;337(8735):215-8.

Grant JM. Induction of labour confers benefits in prolonged pregnancy. [Review]. Br J Obstet Gynaecol 1994;101(2):99-102.

Grant JM. Induction of labour confers benefits in prolonged pregnancy. Br J Obstet Gynaecol 1995;102:80-1.

Graves GR, Baskett TF, Gray JH, et al. The effect of vaginal administration of various doses of prostaglandin E2 gel on cervical ripening and induction of labor. Am J Obstet Gynecol 1985;151(2):178-81.

Gregory KD, Henry OA, Ramicone E, et al. Maternal and infant complications in high and normal weight infants by method of delivery. Obstet Gynecol 1998;92(4 Pt 1):507-13.

Griffiths M. Stillbirths and rate of neonatal deaths in 76,761 postterm pregnancies in Sweden, 1982-1991; a register study. Acta Obstet Gynecol Scand 1998;77(5):583-4.

Grubb DK, Rabello YA, Paul RH. Post-term pregnancy: fetal death rate with antepartum surveillance. [Review]. Obstet Gynecol 1992;79(6):1024-6.

Grünberger W, Spona J. The effect of pericervical PGEinf 2 instillation on levels of maternal serum 13,14-dihydro-15-keto-PGF(2alpha) and progesterone. Arch Gynecol 1986;239(2):93-9.

Guidetti DA, Divon MY, Cavalieri RL, et al. Fetal umbilical artery flow velocimetry in postdate pregnancies. Am J Obstet Gynecol 1987;157(6):1521-3.

Guidetti DA, Divon MY, Langer O. Postdate fetal surveillance: is 41 weeks too early? Am J Obstet Gynecol 1989;161(1):91-3.

Gunderson E. Post-term pregnancy: fetal death rate with antepartum surveillance. Obstet Gynecol 1992;80(4):729-30.

Gupta R, Vasishta K, Sawhney H, et al. Safety and efficacy of stripping of membranes at term. Int J Gynaecol Obstet 1998;60(2):115-21.

Gyetvai K, Hannah ME, Hodnett ED, et al. Tocolytics for preterm labor: A systematic review. Obstet Gynecol 1999;94(5 SUPPL. 1):869-77.

Hagglund L, Christensen KK, Christensen P, et al. Risk factors in cesarean section infection. Obstet Gynecol 1983;62(2):145-50.

Halta VE. Calendars, clocks and choices: prolonged pregnancy [discussion appears in Midwifery Today Childbirth Educ 1996;38:38]. Midwifery Today Childbirth Educ 1996;38:36-7.

Halta VE. Calendars, clocks and choices. Midwifery Today Childbirth Educ 1996;(38):36-8.

Hanley ML, Vintzileos AM. Biophysical testing in premature rupture of the membranes. Semin Perinatol 1996;20(5):418-25.

Hann L, McArdle C, Sachs B. Sonographic biophysical profile in the postdate pregnancy. J Ultrasound Med 1987;6(4):191-5.

Hannah M. Management of post-term pregnancy. In: The Society of Obstetricians and Gynaecologists of Canada: Policy Statement, Committee Opinion, Clinical Practice Guidelines. Ottawa, ON : SOGC, 1996. 10 P.

Hannah M. Post-term pregnancy. In: The Society of Obstetricians and Gynaecologists of Canada: Policy Statement, Committee Opinion, Clinical Practice Guidelines. Ottawa, ON : SOGC, 1997. 7 P.

Hannah ME. Which tests of postterm fetal well-being work? Birth 1990;17(4):228-9.

Hannah ME, Hannah WJ, Caritis SN, et al. Comment on the effectiveness of induction of labor for postterm pregnancy (3). Am J Obstet Gynecol 1995;172(1):240-1.

Hannah ME, Hannah WJ, Hellmann J. Efficacy of induced labor vs serial antenatal monitoring in postterm pregnancy. J Am Osteopath Assoc 1992;92(8):982, 989.

Hannah ME, Hannah WJ, Hellmann J, et al. Induction of labor as compared with serial antenatal monitoring in post-term pregnancy. A randomized controlled trial. The Canadian Multicenter Post-term Pregnancy Trial Group [published erratum appears in N Engl J Med 1992 Jul 30;327(5):368]. N Engl J Med 1992;326(24):1587-92.

Hannah ME, Hannah WJ, Willan A. Comment on the effectiveness of induction of labor for postterm pregnancy. Am J Obstet Gynecol 1995;172(1 Pt 1):240-1.

Hannah ME, Huh C, Hewson SA, et al. Postterm pregnancy: putting the merits of a policy of induction of labor into perspective. Birth 1996;23(1):13-9.

Haque MN. Prolonged pregnancy: Is induction of labour still necessary? J Obstet Gynaecol 1989;9(3):196-8.

Harding K, Evans S, Newnham J. Screening for the small fetus: a study of the relative efficacies of ultrasound biometry and symphysiofundal height. Aust N Z J Obstet Gynaecol 1995;35(2):160-4.

Harman JH, Kim A. Current trends in cervical ripening and labor induction [published erratum appears in Am Fam Physician 1999 Nov 15;60(8):2238]. [Review]. Am Fam Physician 1999;60(2):477-84.

Harris B, Eden RD, Garite TJ, et al. Prolonged pregnancy. Part I: Identifying the patient at risk. Female Patient Pract Obstet Gynecol Med 1991;16(2):43-6, 48, 50.

Harris B, Eden RD, Garite TJ, et al. Prolonged Pregnancy: Part II: Monitoring and intervention. Female Patient Total Health Care Women 1991;16(3):47-9, 53, 56-57, 60.

Harris BA, Huddleston JF, Sutliff G, et al. The unfavorable cervix in prolonged pregnancy. Obstet Gynecol 1983;62(2):171-4.

Harrison JM. Clinical. The initiation of labour: physiological mechanisms. Br J Midwifery 2000;8(5):281-6.

Hart G. Far from settled... the issue of relative risks of "postdates pregnancy" is far from settled. Midwifery Today Childbirth Educ 1996;38:38.

Hauth JC, Goodman MT, Gilstrap LC 3d, et al. Post-term pregnancy. I. Obstet Gynecol 1980;56(4):467-70.

Hauth JC, Hankins GD, Gilstrap LC 3d, et al. Uterine contraction pressures with oxytocin induction/augmentation. Obstet Gynecol 1986;68(3):305-9.

Hedén L, Ingemarsson I, Ahlström H, et al. Induction of labor versus conservative management in prolonged pregnancy: Controlled study. Int J Feto-Matern Med 1991;4(4):231-6.

Hendricksen A. Prolonged pregnancy. A literature review. [Review]. J Nurse Midwifery 1985;30(1):33-42.

Henriksen TB, Wilcox AJ, Hedegaard M, et al. Bias in studies of preterm and postterm delivery due to ultrasound assessment of gestational age. Epidemiology 1995;6(5):533-7.

Herabutya Y, Prasertsawat P, Pokpirom J. A comparison of intravaginal misoprostol and intracervical prostaglandin E2 gel for ripening of unfavorable cervix and labor induction. J Obstet Gynaecol Res 1997;23(4):369-74.

Herabutya Y, Prasertsawat PO, Tongyai T, et al. Prolonged pregnancy: the management dilemma. Int J Gynaecol Obstet 1992;37(4):253-8.

Herbst A, Wolner-Hanssen P, Ingemarsson I. Risk factors for acidemia at birth. Obstet Gynecol 1997;90(1):125-30.

Hilder L, Costeloe K, Thilaganathan B. Prolonged pregnancy: evaluating gestation-specific risks of fetal and infant mortality. Br J Obstet Gynaecol 1998;105(2):169-73.

Hill WC, Fleming AD, Martin RW, et al. Home uterine activity monitoring is associated with a reduction in preterm birth. Obstet Gynecol 1990;76(1 SUPPL.):13S-8S.

Hobart JM. Post-term pregnancy. [Review]. Clin Perinatol 1989;16(4):909-15.

Hofmeyr GJ, Alfirevic Z, Kelly T, et al. Methods for cervical ripening and labour induction in late pregnancy (Cochrane Protocol). In: The Cochrane Library, Issue 3, 2000. Oxford: Update Software.

Hofmeyr GJ, Gulmezoglu AM. Vaginal misoprostol for cervical ripening and labour induction in late pregnancy (Cochrane Review). In: The Cochrane Library, Issue 1, 2001. Oxford: Update Software.

Holcomb WL, Smeltzer JS. Cervical effacement: variation in belief among clinicians. Obstet Gynecol 1991;78(1):43-5.

Holst N, Jenssen TG, Burhol PG, et al. Gastrointestinal regulatory peptides during oxytocin infusion in post-term pregnancies. Acta Physiol Scand 1988;132(1):23-7.

Hourvitz A, Alcalay M, Korach J, et al. A prospective study of high- versus low-dose oxytocin for induction of labor. Acta Obstet Gynecol Scand 1996;75(7):636-41.

Howarth GR, Funk M, Steytler P, et al. A randomised controlled trial comparing vaginally administered misoprostol to vaginal dinoprostone gel in labour induction. J Obstet Gynaecol 1996;16:474-8.

Hsieh TT, Hung TH, Chen KC, et al. Perinatal outcome of oligohydramnios without associated premature rupture of membranes and fetal anomalies. Gynecol Obstet Invest 1998;45(4):232-6.

Huang DY, Usher RH, Kramer MS, et al. Determinants of unexplained antepartum fetal deaths. Obstet Gynecol 2000;95(2):215-21.

Hughes EG, Kelly AJ, Kavanagh J. Dinoprostone vaginal insert for cervical ripening and labor induction: a meta-analysis. Obstet Gynecol 2001;97(5):847-55.

Hulet B, Platt LD. Sinusoidal heart rate pattern and face presentation in a fetus from a postterm pregnancy. A case report. J Reprod Med 1987;32(3):211-2.

Hunter IW. Advanced extra-uterine pregnancy—two cases and a review of the literature. Ir Med J 1983;76(2):82-3.

Hutchon DJ. Gestational age and induction of labour for prolonged pregnancy. Br J Obstet Gynaecol 1998;105(2):247-8.

Hutson JM, Petrie RH. Possible limitations of fetal monitoring. Clin Obstet Gynecol 1986;29(1):104-13.

Idrisa A, Obisesan KA, Adeleye JA. Fetal membrane sweeping for stimulation of labour in prolonged pregnancy: A controlled study. J Obstet Gynaecol 1993;13(4):235-7.

Iffy L, Apuzzio JJ, Mitra S, et al. Rates of cesarean section and perinatal outcome. Perinatal mortality. Acta Obstet Gynecol Scand 1994;73(3):225-30.

Imai M, Tani A, Saito M, et al. Significance of fetal fibronectin and cytokine measurement in the cervicovaginal secretions of women at term in predicting term labor and post-term pregnancy. Eur J Obstet Gynecol Reprod Biol 2001;97:53-8.

Ingardia CJ. Antepartum testing in postterm pregnancy. Am J Obstet Gynecol 1982;143(1):113-4.

Ingemarsson I, Kallen K. Stillbirths and rate of neonatal deaths in 76,761 postterm pregnancies in Sweden, 1982-1991: a register study. [Review]. Acta Obstet Gynecol Scand 1997;76(7):658-62.

Irion O, Boulvain M. Induction of labour for suspected fetal macrosomia (Cochrane Review). In: The Cochrane Library, Issue 4, 2000. Oxford: Update Software.

Ismail AA, Khowesah MM, Shaala SA, et al. Induction of labor by oral prostaglandin E2 in protracted pregnancy. Int J Gynaecol Obstet 1989;29(4):325-8.

Jacobsen G. Prediction of fetal growth deviations by use of symphysis-fundus height measurements. Internat J Tech Assess Health Care 1992;8(Suppl 1):152-9.

Jarvelin MR, Hartikainen-Sorri AL, Rantakallio P. Labour induction policy in hospitals of different levels of specialisation. Br J Obstet Gynaecol 1993;100(4):310-5.

Jazayeri A, Heffron JA, Phillips R, et al. Macrosomia prediction using ultrasound fetal abdominal circumference of 35 centimeters or more. Obstet Gynecol 1999;93(4):523-6.

Jazayeri A, Tsibris JCM, Spellacy WN. Elevated umbilical cord plasma erythropoietin levels in prolonged pregnancies. Obstet Gynecol 1998;92(1):61-3.

Jocums SB, Berg CJ, Entman SS, et al. Postdelivery mortality in Tennessee, 1989-1991. Obstet Gynecol 1998;91(5 I):766-70.

Johanson RB, Menon BKV. Vacuum extraction vs forcepts for assisted vaginal delivery (Cochrane Review). In: The Cochrane Library, Issue 2, 2001. Oxford: Update Software.

Johnson JM, Harman CR, Lange IR, et al. Biophysical profile scoring in the management of the postterm pregnancy: an analysis of 307 patients. Am J Obstet Gynecol 1986;154(2):269-73.

Johnson JW, Longmate JA, Frentzen B. Excessive maternal weight and pregnancy outcome. Am J Obstet Gynecol 1992;167(2):353-70.

Johnson TR, Paine LL, Strobino DM, et al. Population differences affect the interpretation of fetal nonstress test results. Am J Obstet Gynecol 1998;179(3 Pt 1):779-83.

Johnstone MJ. Prolonged pregnancy: is induction of labour indicated? A prospective study. Br J Obstet Gynaecol 1982;89(9):778-9.

Johnstone MJ, Downie G, Gaskell A. Induction of labor for postmaturity. Lancet 1982;2(8296):496.

Jones DC, Hayslett JP. Outcome of pregnancy in women with moderate or severe renal insufficiency. N Engl J Med 1996;335(4):226-32.

Jongsma HW, Nijhuis JG. Critical analysis of the validity of electronic fetal monitoring. J Perinat Med 1991;19(1-2):33-7.

Jorge CS, Artal R, Paul RH, et al. Antepartum fetal surveillance in diabetic pregnant patients. Am J Obstet Gynecol 1981;141(6):641-5.

Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. BMJ 2001;323:42-6.

Jüni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282(11):1054-60.

Kadanali S, Küçüközkan T, Zor N, et al. Comparison of labor induction with misoprostol vs. oxytocin/prostaglandin E2 in term pregnancy. Int J Gynaecol Obstet 1996;55(2):99-104.

Kadar N, Tapp A, Wong A. The influence of nipple

stimulation at term on the duration of pregnancy. J Perinatol 1990;10(2):164-6.

Kanakura Y, Kometani K, Nagata T, et al. Uterine reconstruction for complete septate uterus. Keio J Med 1999;48(4):201-3.

Kaplan B, Goldman GA, Peled Y, et al. The outcome of post-term pregnancy. A comparative study. J Perinat Med 1995;23(3):183-9.

Karande VC, Deshmukh MA, Virkud A. Management of post-term pregnancy. J Postgrad Med 1985;31(2):98-101.

Kassis A, Mazor M, Leiberman JR, et al. Management of post-date pregnancy: a case control study. Isr J Med Sci 1991;27(2):82-6.

Katz VL, Farmer RM, Dean CA, et al. Use of misoprostol for cervical ripening. Southern Med J 2000;93(9):881-4.

Katz Z, Yemini M, Lancet M, et al. Non-aggressive management of post-date pregnancies. Eur J Obstet Gynecol Reprod Biol 1983;15(2):71-9.

Kaufmann RC, McBride P, Amankwah KS, et al. The effect of minor degrees of glucose intolerance on the incidence of neonatal macrosomia. Obstet Gynecol 1992;80(1):97-101.

Keirse MJ. Amniotomy or oxytocin for induction of labor. Re-analysis of a randomized controlled trial. Acta Obstet Gynecol Scand 1988;67(8):731-5.

Keirse MJ. Roundtable: postterm pregnancy. Part II. In the final analysis. Birth 1991;18(2):114-5.

Keirse MJ. Prostaglandins in preinduction cervical ripening. Meta-analysis of worldwide clinical experience. J Reprod Med 1993;38(1 Suppl):89-100.

Keirse MJN. Postterm pregnancy: new lessons from an unresolved debate. Birth 1993;20(2):102-5.

Keirse MJNC. New perspectives for the effective treatment of preterm labor. Am J Obstet Gynecol 1995;173(2):618-28.

Keith RD, Beckley S, Garibaldi JM, et al. A multicentre comparative study of 17 experts and an intelligent computer system for managing labour using the cardiotocogram. Br J Obstet Gynaecol 1995;102(9):688-700.

Kelly AJ, Kavanagh J, Thomas J. Castor oil, bath and/or enema for cervical priming and induction of labour (Cochrane Review). In: The Cochrane Library, Issue 2, 2001. Oxford: Update Software.

Kelly AJ, Kavanagh J, Thomas J. Relaxin for cervical ripening and induction of labour (Cochrane Review). In: The Cochrane Library, Issue 2, 2001. Oxford: Update Software.

Kelly AJ, Kavanagh J, Thomas J. Vaginal prostaglandin (PGE2 and PGF2a) for induction of labour at term (Cochrane Review). In: The Cochrane Library, Issue 2, 2001. Oxford: Update Software.

Kemp B, Winkler M, Rath W. Induction of labor by prostaglandin E2 in relation to the Bishop score. Int J Gynecol Obstet 2000;71:13-7.

Keng MTY, Eng BTS. Oral contraceptives postdatism as a complication. Med J Malaysia 1982;37(4):338-43.

Kesmodel U, Olsen SF. Smoking habits among pregnant Danish women: reliability of information recorded after delivery. J Epidemiol Community Health 1999;53(4):239-42.

Khan KK, Dinnes J, Kleijnen J. Systematic reviews to evaluate diagnostic tests. Eur J Obstet Gynecol Reprod Biol 2001;95:6-11.

Khouzami VA, Johnson JW, Hernandez E, et al. Urinary estrogens in postterm pregnancy. Am J Obstet Gynecol 1981;141(2):205-11.

Kidd LC, Patel NB, Smith R. Non-stress antenatal cardiotocography—a prospective randomized clinical trial. Br J Obstet Gynaecol 1985;92(11):1156-9.

Kilpatrick SJ, Safford KL. Maternal hydration increases amniotic fluid index in women with normal amniotic fluid. Obstet Gynecol 1993;81(1):49-52.

Knox GE, Huddleston JF, Flowers CE. Management of prolonged pregnancy: results of a prospective randomized trial. Am J Obstet Gynecol 1979;134(4):376-84.

Kochenour NK. Estrogen assay during pregnancy. [Review]. Clin Obstet Gynecol 1982;25(4):659-72.

Kochenour NK. Other causes of fetal death. [Review]. Clin Obstet Gynecol 1987;30(2):312-21.

Kolderup LB, Laros RK, Musci TJ. Incidence of persistent birth injury in macrosomic infants: association with mode of delivery. Am J Obstet Gynecol 1997;177(1):37-41.

Konje JC, Obisesan KA, Ladipo OA. Obstructed labor in Ibadan. Int J Gynecol Obstet 1992;39(1):17-21.

Kramer M, et al. The validity of gestational age estimation by menstrual dating in term, preterm, and postterm estations. JAMA 1988;260(22):3306-8.

Kramer RL, Gilson GJ, Morrison DS, et al. A randomized trial of misoprostol and oxytocin for induction of labor: safety and efficacy. Obstet Gynecol 1997;89(3):387-91.

Krammer J, O'Brien WF. Mechanical methods of cervical ripening. [Review]. Clin Obstet Gynecol 1995;38(2):280-6.

Krammer J, Williams MC, Sawai SK, et al. Pre-induction cervical ripening: a randomized comparison of two methods. Obstet Gynecol 1995;85(4):614-8.

Kulkarni SK, Matadial L. Prolonged pregnancy—a rational approach to management. West Indian Med J 1986;35(4):314-7.

Kumari S, Jain S, Pruthi PK, et al. Perinatal risks in postdated pregnancy. Indian Pediatr 1984;21(1):21-7.

Lagrew DC, Freeman RK. Management of postdate pregnancy. Am J Obstet Gynecol 1986;154(1):8-13.

Lahteenmaki P, Rapeli T, Kaariainen M, et al. Late postcoital treatment against pregnancy with antiprogesterone RU 486. Fertil Steril 1988;50(1):36-8.

Landon MB, Gabbe SG. Antepartum fetal surveillance in gestational diabetes mellitus. Diabetes 1985;34(SUPPL. 2):50-4.

Lang J, Lieberman E. Prolonged pregnancy: The management debate. BMJ 1988;297(6650):715-7.

Langer O, Vega-Rich M, Cohen W. Terminal pattern: Characteristics and management. Am J Perinatol 1985;2(4):300-4.

Lazor LZ, Philipson EH, Ingardia CJ, et al. A randomized comparison of 15- and 40-minute dosing protocols for labor augmentation and induction. Obstet Gynecol 1993;82(6):1009-12.

Leaphart WL, Meyer MC, Capeless EL. Labor induction with a prenatal diagnosis of fetal macrosomia. J Matern Fetal Med 1997;6(2):99-102.

Ledger WJ. Identification of the high risk mother and fetus--does it work? [Review]. Clin Perinatol 1980;7(1):125-34.

Lee HY. A randomised double-blind study of vaginal misoprostol vs dinoprostone for cervical ripening and labour induction in prolonged pregnancy. Singapore Med J 1997;38(7):292-4.

Lee NC. Postmaturity and induction of labour. S Afr Med J 1986;70(11):647.

LeFevre ML, Bain RP, Ewigman BG, et al. A randomized trial of prenatal ultrasonographic screening: impact on maternal management and outcome. RADIUS (Routine Antenatal Diagnostic Imaging with Ultrasound) Study Group. Am J Obstet Gynecol 1993;169(3):483-9.

Leiman G, Harrison NA, Rubin A. Pregnancy following conization of the cervix: complications related to cone size. Am J Obstet Gynecol 1980;136(1):14-8.

Lemancewicz A, Urban R, Skotnicki MZ, et al. Uterine and fetal Doppler flow changes after misoprostol and oxytocin therapy for induction of labor in post-term pregnancies. Int J Gynaecol Obstet 1999;67(3):139-45.

Lemons JA, Vargas P, Delaney JJ. Infant of the diabetic mother: Review of 225 cases. Obstet Gynecol 1981;57(2):187-92.

Lennox CE, Patel NB. Early versus late induction of labour in post-term pregnancy. BMJ 1987;294(6588):1689.

Leppert PC. Anatomy and physiology of cervical ripening. Clin Obstet Gynecol 1995;38(2):267-79.

Leslie GI, Gallery EDM, Arnold JD, et al. Neonatal outcome in a randomized, controlled trial of low-dose aspirin in high-risk pregnancies. J Paediatr Child Health 1995;31(6):549-52.

Leveno KJ, Quirk JG, Cunningham FG, et al. Prolonged pregnancy. I. Observations concerning the

causes of fetal distress. Am J Obstet Gynecol 1984;150(5 Pt 1):465-73.

Levine AB, Lockwood CJ, Brown B, et al. Sonographic diagnosis of the large for gestational age fetus at term: does it make a difference? Obstet Gynecol 1992;79(1):55-8.

Lewis R, Mercer BM, Salama M, et al. Oral terbutaline after parenteral tocolysis: A randomized, double- blind, placebo-controlled trial. Am J Obstet Gynecol 1996;175(4 I):834-7.

Librizzi RJ. Antepartum fetal monitoring: nonstress testing in the evaluation of uteroplacental insufficiency. J Am Osteopath Assoc 1980;79(10):636-40.

Lidegaard O, Bottcher LM, Weber T. Description, evaluation and clinical decision making according to various fetal heart rate patterns. Inter-observer and regional variability. Acta Obstet Gynecol Scand 1992;71(1):48-53.

Lien JM, Morgan MA, Garite TJ, et al. Antepartum cervical ripening: applying prostaglandin E2 gel in conjunction with scheduled nonstress tests in postdate pregnancies. Am J Obstet Gynecol 1998;179(2):453-8.

Lim BH, Mahmood TA, Smith NC, et al. A prospective comparative study of transvaginal ultrasonography and digital examination for cervical assessment in the third trimester of pregnancy. J Clin Ultrasound 1992;20(9):599-603.

Lindhard A, Nielsen PV, Mouritsen LA, et al. The implications of introducing the symphyseal-fundal height-measurement. A prospective randomized controlled trial. Br J Obstet Gynaecol 1990;97(8):675-80.

Lipscomb KR, Gregory K, Shaw K. The outcome of macrosomic infants weighing at least 4500 grams: Los Angeles County + University of Southern California experience. Obstet Gynecol 1995;85(4):558-64.

Losh DP, Duhring JL. Management of the postdates pregnancy. Am Fam Physician 1987;36(2):184-94.

Luckas M, Bricker L. Intravenous prostaglandin for induction of labour (Cochrane Review). In: The Cochrane Library, Issue 1, 2001. Oxford: Update Software.

Luckas M, Buckett W, Alfirevic Z. Comparison of outcomes in uncomplicated term and post-term pregnancy following spontaneous labor. J Perinat Med 1998;26(6):475-9.

Lumbiganon P, Laopaiboon M, Kuchaisit C, et al. Oral prostaglandins (excluding misoprostol) for cervical ripening and labour induction when the baby is alive (Cochrane Review). In: The Cochrane Library, Issue 3, 2000. Oxford: Update Software.

Lumley J, Lester A, Anderson I, et al. A randomized trial of weekly cardiotocography in high-risk obstetric patients. Br J Obstet Gynaecol 1983;90(11):1018-26.

Lurie S, Matzkel A, Weissman A, et al. Outcome of pregnancy in class A1 and A2 gestational diabetic patients delivered beyond 40 weeks' gestation. Am J Perinatol 1992;9(5-6):484-8.

Lurie S, Zalel Y, Hagay ZJ. The evaluation of accelerated fetal growth. [Review]. Curr Opin Obstet Gynecol 1995;7(6):477-81.

Lydon-Rochelle M, Holt VL, Easterling TR, et al. Risk of uterine rupture during labor among women with a prior cesarean delivery. N Engl J Med 2001;345(1):3-8.

Maccato M, McLean W, Riddle G, et al. Isolation of Kingella denitrificans from amniotic fluid in a woman with chorioamnionitis. A case report. J Reprod Med 1991;36(9):685-7.

Macer JA, Macer CL, Chan LS. Elective induction versus spontaneous labor: a retrospective study of complications and outcome. Am J Obstet Gynecol 1992;166(6 Pt 1):1690-6.

MacGillivray I. Twins and other multiple deliveries. [Review]. Clin Obstet Gynaecol 1980;7(3):581-600.

MacKenzie IZ, Burns E. Randomised trial of one versus two doses of prostaglandin E2 for induction of labour: 1. Clinical outcome. Br J Obstet Gynaecol 1997;104(9):1062-7.

Macnaughton MC, Chalmers IG, Dubowitz V, et al. Final report of the Medical Research Council/Royal College of Obstetricians and Gynaecologists Multicentre Randomised Trial of Cervical Cerclage. Br J Obstet Gynaecol 1993;100(6):516-23.

Magann EF, Chauhan SP, Nevils BG, et al.

Management of pregnancies beyond forty-one weeks' gestation with an unfavorable cervix. Am J Obstet Gynecol 1998;178(6):1279-87.

Magann EF, McNamara MF, Whitworth NS, et al. Can we decrease postdatism in women with an unfavorable cervix and a negative fetal fibronectin test result at term by serial membrane sweeping? Am J Obstet Gynecol 1998;179(4):890-4.

Mahmood TA. A prospective comparative study on the use of prostaglandin E2 gel (2 mg) and prostaglandin E2 tablet (3 mg) for the induction of labour in primigravid women with unfavorable cervices. Eur J Obstet Gynecol Reprod Biol 1989;33(2):169-75.

Mancuso S, Ferrazzani S, De Carolis S, et al. Term and postterm low-risk pregnancies: management schemes for the reduction of high rates of cesarean section. Minerva Ginecol 1996;48(3):95-8.

Mandelbrot L, Dommergues M, Dumez Y. Prepartum transabdominal amnio-infusion for severe oligohydramnios. Acta Obstet Gynecol Scand 1992;71(2):124-5.

Mandelbrot L, Verspyck E, Dommergues M, et al. Transabdominal amnioinfusion for the management of nonlaboring postdates with severe oligohydramnios. Fetal Diagn Ther 1993;8(6):412-7.

Mandouvalos H, Alkalai BK, Metallinos C, et al. Considerations on the etiology of postponed labour. Ther Hung 1982;30(2):83-7.

Mandruzzato G, Meir YJ, D'Ottavio G, et al. Computerised evaluation of fetal heart rate in post-term fetuses: long term variation. Br J Obstet Gynaecol 1998;105(3):356-9.

Manning FA. The use of sonography in the evaluation of the high-risk pregnancy. Radiol Clin North Am 1990;28(1):205-16.

Mannino F. Neonatal complications of postterm gestation. J Reprod Med Obstet Gynecol 1988;33(3):271-6.

Manolitsas T, Wein P, Beischer NA, et al. Value of cardiotocography in women with antepartum haemorrhage - Is it too late for Caesarean section when the cardiotocograph shows ominous features? Aust N Z J Obstet Gynaecol 1994;34(4):403-8.

Manor M, Blickstein I, Ben-Arie A, et al. Case series of labor induction in twin gestations with an intrauterine balloon catheter. Gynecol Obstet Invest 1999;47(4):244-6.

Martin JN, Perry KG, Roberts WE, et al. Plasma exchange for preeclampsia: II. Unsuccessful antepartum utilization for severe preeclampsia with or without HELLP syndrome. J Clin Apheresis 1994;9(3):155-61.

Martin JN, Sessums JK, Howard P, et al. Alternative approaches to the management of gravidas with prolonged-postterm-postdate pregnancies. J Miss State Med Assoc 1989;30(4):105-11.

Mathews TJ. Trends in stimulation and induction of labor 1989-1995. Stat Bull Metrop Insur Co 1997;78(4):20-6.

Matijevic R. Outcome of post-term pregnancy: a matched-pair case-control study. Croat Med J 1998;39(4):430-4.

Maymon R, Shulman A, Pomeranz M, et al. Uterine rupture at term pregnancy with the use of intracervical prostaglandin E2 gel for induction of labor. Am J Obstet Gynecol 1991;165(2):368-70.

McColgin SW, Hampton HL, McCaul JF, et al. Stripping membranes at term: can it safely reduce the incidence of post-term pregnancies? Obstet Gynecol 1990;76(4):678-80.

McColgin SW, Patrissi GA, Morrison JC. Stripping the fetal membranes at term. Is the procedure safe and efficacious? Journal of Reproductive Medicine 1990;35(8):811-4.

McLaren M, Greer IA, Smith JR, et al. Maternal plasma bicycling PGE2 levels following vaginal administration of prostaglandin E2 pessaries in full term pregnancies. Prog Clin Biol Res 1987;242:199-203.

McLean FH, Boyd ME, Usher RH, et al. Postterm infants: too big or too small? Am J Obstet Gynecol 1991;164(2):619-24.

McMahon MJ, Kuller JA, Yankowitz J. Assessment of the post-term pregnancy. Am Fam Physician 1996;54(2):631-6, 641-2.

McNally OM, Turner MJ. Induction of labour after 1 previous Caesarean section. Aust N Z J Obstet

Gynaecol 1999;39(4):425-9.

Megafu U, Ozumba BC. Obstetric complications of macrosomic babies in African women. Int J Gynaecol Obstet 1988;26(2):197-202.

Mehl-Madrona L, Madrona MM. Physician- and midwife-attended home births. Effects of breech, twin, and post-dates outcome data on mortality rates. J Nurse-Midwifery 1997;42(2):91-8.

Mercer BM, McNanley T, O'Brien JM, et al. Early versus late amniotomy for labor induction: a randomized trial. Am J Obstet Gynecol 1995;173(4):1321-5.

Merrill PA, Porto M, Lovett SM, et al. Evaluation of the nonreactive positive contraction stress test prior to 32 weeks: The role of the biophysical profile. Am J Perinatol 1995;12(4):229-31.

Milchev N, Pehlivanov B, Paskaleva V, et al. Prostaglandin E2 in preinduction cervical ripening in postdate pregnancy. Folia Med (Plovdiv) 1999;41(3):81-5.

Miller DD. Tocolytic therapy for the prevention of preterm labor. J Intraven Nurs 1995;18(5):233-8.

Miller FC, Read JA. Intrapartum assessment of the postdate fetus. Am J Obstet Gynecol 1981;141(5):516-20.

Minaretzis D, Tsionu C, Papageorgiou I, et al. Intracervical prostaglandin E2 gel for cervical ripening and labor induction: what is the appropriate dose? Gynecol Obstet Invest 1993;35(1):34-7.

Minchom P, Niswander K, Chalmers I, et al. Antecedents and outcome of very early neonatal seizures in infants born at or after term. Br J Obstet Gynaecol 1987;94(5):431-9.

Misra M, Vavre S. Labour induction with intracervical prostaglandin Einf 2 gel and intravenous oxytocin in women with a very unfavourable cervix. Aust N Z J Obstet Gynaecol 1994;34(5):511-5.

Mittendorf R, Williams MA, Berkey CS, et al. The length of uncomplicated human gestation. Obstet Gynecol 1990;75(6):929-32.

Miyakawa I, Ichimaru S, Tayama C, et al. Effect of ACTH infusion on induction of uterine contraction in prolonged human pregnancy in association with endocrine function. Int J Gynecol Obstet 1981;19(2):89-96.

Miyazaki FS, Miyazaki BA. False reactive nonstress tests in postterm pregnancies. Am J Obstet Gynecol 1981;140(3):269-76.

Mizutani S, Itakura A, Kurauchi O, et al. Maternal serum oxytocinase activities in patients with poor obstetric dates. Med Sci Res 1994;22(11):771-5.

Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. Internat J Tech Assess Health Care 1996;12(2):195-208.

Mohide PT. Randomised comparison of early versus late induction of labor in post-term pregnancy. BMJ 1987;295(6594):388-9.

Moldin PG, Sundell G. Induction of labour: a randomised clinical trial of amniotomy versus amniotomy with oxytocin infusion. Br J Obstet Gynaecol 1996;103(4):306-12.

Monaghan J, O'Herlihy C, Boylan P. Ultrasound placental grading and amniotic fluid quantitation in prolonged pregnancy. Obstet Gynecol 1987;70(3 Pt 1):349-52.

Mongan PF, Fadel HE. Postdatism. Am Fam Physician 1981;23(3):208-12.

Monincx WM, Zondervan HA, Birnie E, et al. High risk pregnancy monitored antenatally at home. Eur J Obstet Gynecol Reprod Biol 1997;75(2):147-53.

Montan S, Malcus P. Amniotic fluid index in prolonged pregnancy: A cohort study. J Matern Fetal Invest 1995;5(1):4-7.

Moran C, Carranza-Lira S, Ochoa R, et al. Gastrin levels in mothers and neonates at delivery in various perinatal conditions. Acta Obstet Gynecol Scand 1996;75(7):608-11.

Moran DJ, McGarrigle HH, Lachelin GC. Lack of normal increase in saliva estriol/progesterone ratio in women with labor induced at 42 weeks' gestation. Am J Obstet Gynecol 1992;167(6):1563-4.

Mouw RJ, Egberts J, Kragt H, et al. Cervicovaginal fetal fibronectin concentrations: predictive value of impending birth in postterm pregnancies. Eur J Obstet Gynecol Reprod Biol 1998;80(1):67-70.

Mukherji J, Samaddar JC. How safe is caesarean section. J Obstet Gynaecol 1995;21(1):17-21.

Muller PR, Stubbs TM, Laurent SL. A prospective randomized clinical trial comparing two oxytocin induction protocols. Am J Obstet Gynecol 1992;167(2):373-80.

Mundle WR, Young DC. Vaginal misoprostol for induction of labor: a randomized controlled trial. Obstet Gynecol 1996;88(4 Pt 1):521-5.

Murphy PA, Fullerton J. Outcomes of intended home births in nurse-midwifery practice: a prospective descriptive study. [Review]. Obstet Gynecol 1998;92(3):461-70.

Nageotte MP, Freeman RK, Freeman AG, et al. Short-term variability assessment from abdominal electrocardiogram during the antepartum period. Am J Obstet Gynecol 1983;145(5):566-9.

Nagey DA. Post-term pregnancy. Obstet Gynecol 1981;58(1):135-6.

National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. A clinical trial of induction of labor versus expectant management in postterm pregnancy. The National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. Am J Obstet Gynecol 1994;170(3):716-23.

Nationwide Inpatient Sample [NIS]. [electronic database]. Release 6; 1997 data. Agency for Healthcare Research and Quality: Rockville, MD. Available from: http://www.ntis.gov/fcpc/cpn8834.htm. 1997.

Neerhof MG, Cravello C, Haney EI, et al. Timing of labor induction after premature rupture of membranes between 32 and 36 weeks' gestation. Am J Obstet Gynecol 1999;180(2 I):349-52.

Neiger R. Fetal macrosomia in the diabetic patient. Clin Obstet Gynecol 1992;35(1):138-50.

Neilson JP. Symphysis-fundal height measurement in pregnancy (Cochrane Review). In: The Cochrane Library, Issue 3, 2000. Oxford: Update Software.

Neilson JP. Ultrasound for fetal assessment in early pregnancy (Cochrane Review). In: The Cochrane Library, Issue 4, 2000. Oxford: Update Software.

Neilson JP. Mifepristone for induction of labour (Cochrane Review). In: The Cochrane Library, Issue 3, 2001. Oxford: Update Software.

Neilson JP, Alfirevic Z. Doppler ultrasound for fetal assessment in high risk pregnancies (Cochrane Review). In: The Cochrane Library, Issue 3, 2000. Oxford: Update Software.

Neri A, Kaplan B, Rabinerson D, et al. The management of persistent occipito-posterior position. Clin Exp Obstet Gynecol 1995;22(2):126-31.

Newnham JP, Evans SF, Michael CA, et al. Effects of frequent ultrasound during pregnancy: a randomised controlled trial. Lancet 1993;342(8876):887-91.

Nichols CW. Postdate pregnancy: clinical implications... part 2. J Nurse Midwifery 1985;30(5):259-68.

Nichols CW. Postdate pregnancy. Part I. A literature review. [Review]. J Nurse Midwifery 1985;30(4):222-39.

Nocon JJ, McKenzie DK, Thomas LJ, et al. Shoulder dystocia: an analysis of risks and obstetric maneuvers. Am J Obstet Gynecol 1993;168(6 Pt 1):1732-7.

Norman K, Pattinson RC, De SJ, et al. Ampicillin and metronidazole treatment in preterm labour: A multicentre, randomised controlled trial. Br J Obstet Gynaecol 1994;101(5):404-8.

Novakov A, Segedi D, Milasinovic L, et al. Induction of labor by endocervical application of prostaglandins and intravenous infusion of oxytocin in postterm pregnancy. Med Pregl 1998;51(9-10):419-26.

Novy MJ, Walsh SW. Dexamethasone and estradiol treatment in pregnant rhesus macaques: Effects on gestational length, maternal plasma hormones, and fetal growth. Am J Obstet Gynecol 1983;145(8):920-31.

O'Brien JM, Mercer BM, Cleary NT, et al. Efficacy of outpatient induction with low-dose intravaginal prostaglandin E2: a randomized, double-blind, placebo-controlled trial. Am J Obstet Gynecol 1995;173(6):1855-9.

O'Connor R. Induction of labour at term. Evidence on

outcome favours induction. BMJ 1993;306(6889):1413-4.

O'Leary JA, Leonetti HB. Shoulder dystocia: prevention and treatment. Am J Obstet Gynecol 1990;162(1):5-9.

O'Reilly-Green CP, Divon MY. Predictive value of amniotic fluid index for oligohydramnios in patients with prolonged pregnancies. J Matern Fetal Med 1996;5(4):218-26.

O'Reilly-Green CP, Divon MY. Receiver operating characteristic curves of sonographic estimated fetal weight for prediction of macrosomia in prolonged pregnancies. Ultrasound Obstet Gynecol 1997;9(6):403-8.

Ocak V, Sen C, Demirkiran F, et al. FHR monitoring and perinatal mortality in high-risk pregnancies. Eur J Obstet Gynecol Reprod Biol 1992;44(1):59-63.

Ohel G, Rahav D, Rothbart H, et al. Randomised trial of outpatient induction of labor with vaginal PGE2 at 40-41 weeks of gestation versus expectant management. Arch Gynecol Obstet 1996;258(3):109-12.

Ohel G, Yaacobi N, Linder N, et al. Postdate antenatal testing. Int J Gynaecol Obstet 1995;49(2):145-7.

Ojo A, Oronsaye U. Who is the elderly primigravida in Nigeria? Int J Gynecol Obstet 1988;26(1):51-5.

Olofsson P, Saldeen P. The prospects for vaginal delivery in gestations beyond 43 weeks. Acta Obstet Gynecol Scand 1996;75(7):645-50.

Olofsson P, Saldeen P, Marsal K. Association between a low umbilical artery pulsatility index and fetal distress in labor in very prolonged pregnancies. Eur J Obstet Gynecol Reprod Biol 1997;73(1):23-9.

Olsen K. Taken to the limit. Pract Midwife 1999;2(6):46.

Orhue AA. A randomised trial of 45 minutes and 15 minutes incremental oxytocin infusion regimes for the induction of labour in women of high parity. Br J Obstet Gynaecol 1993;100(2):126-9.

Orhue AA. A randomized trial of 30-min and 15-min oxytocin infusion regimen for induction of labor at term in women of low parity. Int J Gynaecol Obstet 1993;40(3):219-25.

Orhue AA. Incremental increases in oxytocin infusion regimens for induction of labor at term in primigravidas: a randomized controlled trial. Obstet Gynecol 1994;83(2):229-33.

Otto C, Platt LD. Fetal growth and development. [Review]. Obstet. Gynecol. Clin. North Am. 1991;18(4):907-31.

Owen J, Henson BV, Hauth JC. A prospective randomized study of saline solution amnioinfusion. Am J Obstet Gynecol 1990;162(5):1146-9.

Owen J, Winkler CL, Harris BA, et al. A randomized, double-blind trial of prostaglandin E2 gel for cervical ripening and meta-analysis. Am J Obstet Gynecol 1991;165(4 Pt 1):991-6.

Oyarzun E, Gomez R, Rioseco A, et al. Antibiotic treatment in preterm labor and intact membranes: A randomized, double-blinded, placebo-controlled trial. J Matern Fetal Med 1998;7(3):105-10.

Papageorgiou I, Tsionou C, Minaretzis D, et al. Labor characteristics of uncomplicated prolonged pregnancies after induction with intracervical prostaglandin E2 gel versus intravenous oxytocin. Gynecol Obstet Invest 1992;34(2):92-6.

Park MI, Hwang JH, Cha KJ, et al. Computerized analysis of fetal heart rate parameters by gestational age. Int J Gynecol Obstet 2001;74:157-64.

Parry E, Parry D, Pattison N. Induction of labour for post term pregnancy: an observational study. Aust N Z J Obstet Gynaecol 1998;38(3):275-80.

Pattison N, McCowan L. Cardiotocography for antepartum fetal assessment (Cochrane Review). In: The Cochrane Library, Issue 1, 2001. Oxford: Update Software.

Pearce JM, Cardozo C. Prolonged pregnancy: the management debate. BMJ 1988;297:715-7.

Peck T. Electronic monitoring evidence of fetal distress in high-risk pregnancies. J Reprod Med 1980;24(3):103-8.

Perkins RP. Sudden fetal death in labor. The significance of antecedent monitoring characteristics and clinical circumstances. J Reprod Med 1980;25(6):309-14.

Persson-Kjerstadius N, Forsgren H, Westgren M. Intrapartum amnioinfusion in women with oligohydramniosis. A prospective randomized trial. Acta Obstet Gynecol Scand 1999;78(2):116-9.

Peterson LP, Kundu N. Endocrine assessment of high-risk pregnancies. Obstet Gynecol Annu 1980; 9:169-94.

Phelan JP. Medical-legal considerations in the postdate pregnancy. Clin Obstet Gynecol 1989;32(2):294-302.

Phelan JP, Platt LD, Yeh SY, et al. The role of ultrasound assessment of amniotic fluid volume in the management of the postdate pregnancy. Am J Obstet Gynecol 1985;151(3):304-8.

Phelan JP, Platt LD, Yeh SY, et al. Continuing role of the nonstress test in the management of postdates pregnancy. Obstet Gynecol 1984;64(5):624-8.

Phelps JY, Higby K, Smyth MH, et al. Accuracy and intraobserver variability of simulated cervical dilatation measurements. Am J Obstet Gynecol 1995;173(3 Pt 1):942-5.

Phillips K, Berry C, Mathers AM. Uterine rupture during second trimester termination of pregnancy using mifepristone and a prostaglandin. Eur J Obstet Gynecol Reprod Biol 1996;65(2):175-6.

Piccolo AR. Clinical management of postdate pregnancy in the small community hospital. J Am Osteopath Assoc 1980;80(4):271-5.

Plaut MM, Schwartz ML, Lubarsky SL. Uterine rupture associated with the use of misoprostol in the gravid patient with a previous cesarean section. Am J Obstet Gynecol 1999;180(6 Pt 1):1535-42.

Pollack RN, Hauer-Pollack G, Divon MY. Macrosomia in postdates pregnancies: the accuracy of routine ultrasonographic screening. Am J Obstet Gynecol 1992;167(1):7-11.

Porreco RP. Postdates management: induction after 41 weeks, with cervical preparation. Birth 1990;17(4):227.

Porto M. The unfavorable cervix: methods of cervical priming. [Review]. Clin Obstet Gynecol 1989;32(2):262-8.

Ramrekersingh-White P, Farkas AG, Chard T, et al. Self-selected expectant management of post dates pregnancy including the use of Doppler ultrasound. J Obstet Gynaecol 1993;13(1):16-9.

Ramsey P, Harris D, Ogburn P, et al. Comparative efficacy of prostaglandin analogues dinoprostone and misoprostol as labor preinduction agents. Am J Obstet Gynecol 1998;178:S94.

Rand L, Robinson JN, Economy KE, et al. Post-term induction of labor revisited. Obstet Gynecol 2000;96(5):779-83.

Rasmussen S, Dalaker K, Nordbo BL, et al. One-stage ultrasound screening in pregnancy. An evaluation. Acta Obstet Gynecol Scand 1990;69(7-8):581-8.

Rayburn W, Gosen R, Ramadei C, et al. Outpatient cervical ripening with prostaglandin E2 gel in uncomplicated postdate pregnancies. Am J Obstet Gynecol 1988;158(6 Pt 1):1417-23.

Rayburn WF, Chang FE. Management of the uncomplicated postdate pregnancy. J Reprod Med 1981;26(2):93-5.

Rayburn WF, Motley ME, Stempel LE, et al. Antepartum prediction of the postmature infant. Obstet Gynecol 1982;60(2):148-53.

Reilly KEH. Induction of labor. Am Fam Physician 1994;49(6):1427-32.

Rekwot PI, Oyedipe EO, Mukasa-Mugerwa E, et al. Fertility in zebu cattle (Bos indicus) after prostaglandin administration and artificial insemination. Vet J 1999;158(1):53-8.

Ribbert LSM, Visser GHA, Mulder EJH, et al. Changes with time in fetal heart rate variation, movement incidences and haemodynamics in intrauterine growth retarded fetuses: A longitudinal approach to the assessment of fetal well being. Early Hum Dev 1993;31(3):195-208.

Ries M, Beinder E, Gruner C, et al. Rapid development of hydrops fetalis in the donor twin following death of the recipient twin in twin-twin transfusion syndrome. J Perinat Med 1999;27(1):68-73.

Rizzo N, Farina A, Santarsiero G, et al. Amniotic fluid index and labor length of pregnancies induced beyond 41 weeks of gestation with unfavorable

cervix. Gynecol Obstet Invest 2000;49(4):244-8.

Roach VJ, Rogers MS. Pregnancy outcome beyond 41 weeks gestation. Int J Gynaecol Obstet 1997;59(1):19-24.

Roberts CL, Algert CS, March LM. Delayed childbearing—are there any risks? Med J Aust 1994;160(9):539-44.

Roberts CL, Taylor L, Henderson-Smart D. Trends in births at and beyond term: evidence of a change? Br J Obstet Gynaecol 1999;106(9):937-42.

Roberts L, Cook E, Beardsworth SA, et al. Prolonged pregnancy: two years experience of offering women conservative management. J R Army Med Corps 1994;140(1):32-6.

Roberts LJ. Induction of labour. BMJ 1993;307(6895):66-7.

Roberts LJ, Young KR. The management of prolonged pregnancy—an analysis of women's attitudes before and after term. Br J Obstet Gynaecol 1991;98(11):1102-6.

Roberts WE, Morrison JC. Has the use of home monitors, fetal fibronectin, and measurement of cervical length helped predict labor and/or prevent preterm delivery in twins? Clin Obstet Gynecol 1998;41(1):95-102.

Rodriguez MH. Ultrasound evaluation of the postdate pregnancy. [Review]. Clin Obstet Gynecol 1989;32(2):257-61.

Romero R, Sibai BM, Sanchez-Ramos L, et al. An oxytocin receptor antagonist (atosiban) in the treatment of preterm labor: A randomized, double-blind, placebo-controlled trial with tocolytic rescue. Am J Obstet Gynecol 2000;182(5):1173-83.

Ron M, Adoni A, Hochner-Celnikier D, et al. The significance of baseline tachycardia in the postterm fetus. Int J Gynaecol Obstet 1980;18(1):76-7.

Rooks JP, Weatherby NL, Ernst EK. The National Birth Center Study. Part III. Intrapartum and immediate postpartum and neonatal complications and transfers, postpartum and neonatal care, outcomes, and client satisfaction. J Nurse Midwifery 1992;37(6):361-97.

Rose VL. ACOG issues report on the management of post-term pregnancy. American College of Obstetricians and Gynecologists. Am Fam Physician 1998;57(7):1686-7.

Rosen MG, Dickinson JC. Management of post-term pregnancy. N Engl J Med 1992;326(24):1628-9.

Rosser J, Anderson T. Sweeping the membranes. Pract Midwife 1998;1(3):28-9.

Rouse DJ, Owen J, Goldenberg RL, et al. The effectiveness and costs of elective cesarean delivery for fetal macrosomia diagnosed by ultrasound. JAMA 1996;276(18):1480-6.

Roussis P, Cox SM, Campbell BA, et al. Survey on the management of post date pregnancy. J Matern Fetal Invest 1993;2(3):155-7.

Royal College of Obstetricians and Gynaecologists. Induction of labour: evidence-based clinical guideline number 9. London: Royal College of Obstetricians and Gynaecologists, 2001.

Rubinstein TH, Schifrin BS. Prolonged labor with persistent occiput-posterior position in postterm pregnancy. J Perinatol 1992;12(2):181-4.

Ryo E, Kozuma S, Sultana J, et al. Fetal size as a determinant of obstetrical outcome of post-term pregnancy. Gynecol Obstet Invest 1999;47(3):172-6.

Sachs BP, Friedman EA. Results of an epidemiologic study of postdate pregnancy. J Reprod Med 1986;31(3):162-6.

Salamalekis E, Vitoratos N, Kassanos D, et al. Sweeping of the membranes versus uterine stimulation by oxytocin in nulliparous women. A randomized controlled trial. Gynecol Obstet Invest 2000;49(4):240-3.

Salamalekis E, Vitoratos N, Loghis C, et al. The predictive value of a nonstress test taken 24 h before delivery in high-risk pregnancies. Int J Gynaecol Obstet 1994;45(2):105-7.

Sanchez-Ramos L, Kaunitz AM, Del Valle GO, et al. Labor induction with the prostaglandin E1 methyl analogue misoprostol versus oxytocin: a randomized trial. Obstet Gynecol 1993;81(3):332-6.

Sandmire HF, DeMott RK. The Green Bay cesarean section study. IV. The physician factor as a determinant of cesarean birth rates for the large fetus.

Am J Obstet Gynecol 1996;174(5):1557-64.

Sarkar PK, Duthie SJ. The clinical significance of reduced amniotic fluid index in post-term pregnancy: a retrospective study. J Obstet Gynaecol 1997;17(3):274-5.

Sarkar PK, Williams J, Duthie J. Fetal macrosomia (1). J Obstet Gynaecol 1994;14(4):293.

Sarno AP, Hinderstein WN, Staiano RA. Fetal macrosomia in a military hospital: Incidence, risk factors, and outcome. Mil Med 1991;156(2):55-8.

Sarno AP, Polzin WJ, Feinstein SJ, et al. Transabdominal amnioinfusion in preterm pregnancies complicated by fetal growth restriction, oligohydramnios and umbilical cord compression. Fetal Diagn Ther 1995;10(6):408-14.

Satin AJ, Hankins GD. Induction of labor in the postdate fetus. [Review]. Clin Obstet Gynecol 1989;32(2):269-77.

Satin AJ, Hankins GD, Yeomans ER. A prospective study of two dosing regimens of oxytocin for the induction of labor in patients with unfavorable cervices. Am J Obstet Gynecol 1991;165(4 Pt 1):980-4.

Satin AJ, Leveno KJ, Sherman ML, et al. High-dose oxytocin: 20- versus 40-minute dosage interval. Obstet Gynecol 1994;83(2):234-8.

Sawai SK, O'Brien WF, Mastrogiannis DS, et al. Patient-administered outpatient intravaginal prostaglandin E2 suppositories in post-date pregnancies: a double-blind, randomized, placebo-controlled study. Obstet Gynecol 1994;84(5):807-10.

Sawai SK, Williams MC, O'Brien WF, et al. Sequential outpatient application of intravaginal prostaglandin E2 gel in the management of postdates pregnancies. Obstet Gynecol 1991;78(1):19-23.

Schan P. Why?... inter-uterine death at 42 weeks. Midwifery Matters 1997;74:12.

Schreyer P, Bar-Natan N, Sherman DJ, et al. Fetal breathing movements before oxytocin induction in prolonged pregnancies. Am J Obstet Gynecol 1991;165(3):577-81.

Schwartz RW. Pregnancy in physicians: characteristics and complications. Obstet Gynecol 1985;66(5):672-6.

Sciscione AC, Nguyen L, Manley J, et al. A randomized comparison of transcervical Foley catheter to intravaginal misoprostol for preinduction cervical ripening. Obstet Gynecol 2001;97:603-7.

Sciscione AC, Nguyen L, Manley JS, et al. Uterine rupture during preinduction cervical ripening with misoprostol in a patient with a previous Caesarean delivery. Aust N Z J Obstet Gynaecol 1998;38(1):96-7.

Scollo P. Epidemiology of cesarean sections: prolonged pregnancy. [Review]. Clin Exp Obstet Gynecol 1999;26(1):22-6.

Semczuk M, Lopucka M. Evaluation of cervix condition according to Bishop score in post term pregnancy. Ann Univ Mariae Curie Sklodowska [Med] 1986;41:125-31.

Setness P. Prolonged pregnancy. When to wait, when to intervene. Postgrad Med 1988;84(3):61-5, 69.

Seyb ST, Berka RJ, Socol ML, et al. Risk of cesarean delivery with elective induction of labor at term in nulliparous women. Obstet Gynecol 1999;94(4):600-7.

Shapiro H, Lyons E. Late maternal age and postdate pregnancy. Am J Obstet Gynecol 1989;160(4):909-12.

Sharma JB, Smith RJ, Wilkin DJ. Induction of labour at term. Women not for waiting. BMJ 1993;306(6889):1413.

Shaw K, Clark SL. Reliability of intrapartum fetal heart rate monitoring in the postterm fetus with meconium passage. Obstet Gynecol 1988;72(6):886-9.

Shaw KJ, Medearis AL, Horenstein J, et al. Selective labor induction in postterm patients. Observations and outcomes. J Reprod Med 1992;37(2):157-61.

Shea KM, Wilcox AJ, Little RE. Postterm delivery: a challenge for epidemiologic research. Epidemiology 1998;9(2):199-204.

Shearer MH, Estes M. A critical review of the recent literature on postterm pregnancy and a look at women's experiences. Birth 1985;12(2):95-111.

Sherer DM, Onyeije CI, Binder D, et al. Uncomplicated baseline fetal tachycardia or bradycardia in postterm pregnancies and perinatal outcome. Am J Perinatol 1998;15(5):335-8.

Sherman DJ, Frenkel E, Pansky M, et al. Balloon cervical ripening with extra-amniotic infusion of saline or prostaglandin E2: a double-blind, randomized controlled study. Obstet Gynecol 2001;97(3):375-80.

Shime J. Influence of prolonged pregnancy on infant development. J Reprod Med Obstet Gynecol 1988;33(3):277-84.

Shime J, Gare DJ, Andrews J, et al. Prolonged pregnancy: surveillance of the fetus and the neonate and the course of labor and delivery. Am J Obstet Gynecol 1984;148(5):547-52.

Shime J, Librach CL, Gare DJ, et al. The influence of prolonged pregnancy on infant development at one and two years of age: a prospective controlled study. Am J Obstet Gynecol 1986;154(2):341-5.

Sholl JS. Abruptio placentae: Clinical management in nonacute cases. Am J Obstet Gynecol 1987;156(1):40-51.

Shy KK. Roundtable: postterm pregnancy. Antenatal testing and candid reassurance... part 2. Birth 1991;18(2):113.

Sibai BM, Taslimi M, Abdella TN, et al. Maternal and perinatal outcome of conservative management of severe preeclampsia in midtrimester. Am J Obstet Gynecol 1985;152(1):32-7.

Silber M, Larsson B, Uvnas-Moberg K. Oxytocin, somatostatin, insulin and gastrin concentrations vis-a-vis late pregnancy, breastfeeding and oral contraceptives. Acta Obstet Gynecol Scand 1991;70(4-5):283-9.

Sims ME, Walther FJ. Neonatal morbidity and mortality and long-term outcome of postdate infants. [Review]. Clin Obstet Gynecol 1989;32(2):285-93.

Skretek M, Bielecki M, Zdanowicz A, et al. Prognostic value of cardiotocographic acoustic test in post-term pregnancies. Ann Med Univ Bialyst Pol 1993;38(1):79-85.

Small ML, Phelan JP, Smith CV, et al. An active management approach to the postdate fetus with a reactive nonstress test and fetal heart rate decelerations. Obstet Gynecol 1987;70(4):636-40.

Smith CA, Crowther CA. Acupuncture for induction of labour (Cochrane Review). In: The Cochrane Library, Issue 1, 2001. Oxford: Update Software.

Smith GCS. Life-table analysis of the risk of perinatal death at term and post term in singleton pregnancies. Am J Obstet Gynecol 2001;184:489-96.

Snyder SG, King EA. Prostaglandin gel for cervical ripening. J Fam Pract 1996;43(1):19-20.

Soper DE, Mayhall CG, Froggatt JW. Characterization and control of intraamniotic infection in an urban teaching hospital. Am J Obstet Gynecol 1996;175(2):304-9.

Spong CY, Beall M, Rodrigues D, et al. An objective definition of shoulder dystocia: prolonged head-to-body delivery intervals and/or the use of ancillary obstetric maneuvers. Obstet Gynecol 1995;86(3):433-6.

Sporken JMJ, Hein PR, Gerritsen SM. First experiences with Swan-Ganz measurements in patients with severe hypertension during pregnancy from an unselected population. Eur J Obstet Gynecol Reprod Biol 1986;21(3):135-41.

Srisomboon J, Piyamongkol W, Aiewsakul P. Comparison of intracervical and intravaginal misoprostol for cervical ripening and labour induction in patients with an unfavourable cervix. J Med Assoc Thai 1997;80(3):189-94.

Steel SA, Pearce JM. Specific therapy in severe fetal intrauterine growth retardation: failure of prostacyclin. J R Soc Med 1988;81(4):214-6.

Steer PJ. Postmaturity—much ado about nothing. Br J Obstet Gynaecol 1986;93(2):105-8.

Stenlund PM, Ekman G, Aedo A-R, et al. Induction of labor with mifepristone: a randomized, double-blind study versus placebo. Acta Obstet Gynecol Scand 1999;78(9):793-8.

Steyn DW, Odendaal HJ. Computerised cardiotocography in a high-risk unit in a developing country—its influence on inter-observer variation and duration of recording. S Afr Med J 1996;86(2):172-5.

Stroup DF, Berlin JA, Morton SC, et al. Meta-

analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283(15):2008-12.

Stubblefield PG, Berek JS. Perinatal mortality in term and post-term births. Obstet Gynecol 1980;56(6):676-82.

Sue-A-Quan AK, Hannah ME, Cohen MM, et al. Effect of labour induction on rates of stillbirth and cesarean section in post-term pregnancies. CMAJ 1999;160(8):1145-9.

Suikkari AM, Jalkanen M, Heiskala H, et al. Prolonged pregnancy: induction or observation. Acta Obstet Gynecol Scand 1983;116(Suppl):58.

Sulik SM, Greenwald JL. Evaluation and management of postdate pregnancy. [Review]. Am Fam Physician 1994;49(5):1177-86, 1191-2.

Summers L. Methods of cervical ripening and labor induction. [Review]. J Nurse Midwifery 1997;42(2):71-85.

Surbek DV, Boesiger H, Hoesli I, et al. A double-blind comparison of the safety and efficacy of intravaginal misoprostol and prostaglandin E2 to induce labor. Am J Obstet Gynecol 1997;177(5):1018-23.

Suzuki S, Otsubo Y, Sawa R, et al. Clinical trial of induction of labor versus expectant management in twin pregnancy. Gynecol Obstet Invest 2000;49(1):24-7.

Tabor B, Anderson J, Stettler B, et al. Misoprostol vs prostaglandin E2 gel for cervical ripening. Am J Obstet Gynecol 1995;172:425.

Tai-Seale M, Rodwin M, Wedig G. Drive-through delivery: where are the "savings"? Med Care Res Rev 1999;56(1):30-46.

Takahashi K, Diamond F, Bieniarz J, et al. Uterine contractility and oxytocin sensitivity in preterm, term, and postterm pregnancy. Am J Obstet Gynecol 1980;136(6):774-9.

Tal Z, Frankel ZN, Ballas S, et al. Breast electrostimulation for the induction of labor. Obstet Gynecol 1988;72(4):671-4.

Tam W-H, Tai SMB, Rogers MS. Prediction of cervical response to prostaglandin Einf 2 using fetal fibronectin. Acta Obstet Gynecol Scand 1999;78(10):861-5.

Tan KH. Fetal manipulation for facilitating tests of fetal wellbeing (Cochrane Protocol). In: The Cochrane Library, Issue 3, 2000. Oxford: Update Software.

Taylor HA. Induction of labor vs. antenatal monitoring in post-term pregnancy. Am Fam Physician 1997;55(7):2428, 2431.

Taylor NF. Review: placental sulphatase deficiency. [Review]. J Inherit Metab Dis 1982;5(3):164-76.

Tchobroutsky C, Vray MM, Altman J-J. Risk/benefit ratio of changing late obstetrical strategies of insulin-dependent diabetic pregnancies. A comparison between 1971-1977 and 1978-1985 periods in 389 pregnancies. Diabetes Metabol 1991;17(2):287-94.

Tessarolo M, Bellino R, Arduino S, et al. Cervical cerclage for the treatment of patients with placenta previa. Clin Exp Obstet Gynecol 1996;23(3):184-7.

Thacker SB, Stroup DF. Continuous electronic heart rate monitoring for fetal assessment during labor (Cochrane Review). In: The Cochrane Library, Issue 1, 2001. Oxford: Update Software.

Thompson G, Newnham JP, Roberman BD, et al. Contraction stress fetal heart rate monitoring at preterm gestational ages. Aust N Z J Obstet Gynaecol 1990;30(2):120-3.

Thornton YS, Yeh SY, Petrie RH. Antepartum fetal heart rate testing and the post-term gestation. J Perinat Med 1982;10(4):196-202.

Todros T, Preve CU, Plazzotta C, et al. Fetal heart rate tracings: observers versus computer assessment. Eur J Obstet Gynecol Reprod Biol 1996;68(1-2):83-6.

Tongsong T, Srisomboon J. Amniotic fluid volume as a predictor of fetal distress in postterm pregnancy. Int J Gynaecol Obstet 1993;40(3):213-7.

Toppozada MK, Anwar MY, Hassan HA, et al. Oral or vaginal misoprostol for induction of labor. Int J Gynaecol Obstet 1997;56(2):135-9.

Toth M, Rehnstrom J, Fuchs AR. Prostaglandins E and F in cervical mucus of pregnant women. Am J

Perinatol 1989;6(2):142-4.

Treacy B, Smith C, Rayburn W. Ultrasound in labor and delivery. [Review]. Obstet Gynecol Surv 1990;45(4):213-9.

Trimmer KJ, Leveno KJ, Peters MT, et al. Observations on the cause of oligohydramnios in prolonged pregnancy. Am J Obstet Gynecol 1990;163(6 I):1900-3.

Trofatter KF. Endocervical prostaglandin E2 gel for preinduction cervical ripening. Clinical trial results. J Reprod Med 1993;38(1 Suppl):78-82.

Tsuei JJ, Lai Y, Sharma SD. The influence of acupuncture stimulation during pregnancy: the induction and inhibition of labor. Obstet Gynecol 1977;50(4):479-8.

Tuffnell DJ, Bryce F, Johnson N, et al. Simulation of cervical changes in labour: reproducibility of expert assessment. Lancet 1989;2(8671):1089-90.

Tuncer R, Erkaya S, Sipahi T, et al. Emergency postpartum hysterectomy. J Gynecol Surg 1995;11(4):209-13.

Tunon K, Eik-Nes SH, Grottum P. Fetal outcome in pregnancies defined as post-term according to the last menstrual period estimate, but not according to the ultrasound estimate. Ultrasound Obstet Gynecol 1999;14(1):12-6.

Turner JE, Burke MS, Porreco RP, et al. Prostaglandin E2 in tylose gel for cervical ripening before induction of labor. J Reprod Med 1987;32(11):815-21.

Turner MJ, Fox R, Gordon H. Induction of labour in primiparae after 41 weeks of pregnancy using vaginal prostaglandins. J Obstet Gynaecol 1988;8(Suppl. 1):S14-S15.

Turnley MJ. Round table: postterm pregnancy. Our moment with Courtney. Birth 1990;17(4):224-6.

Ugwumadu A. Induction of labour confers benefits in prolonged pregnancy. Br J Obstet Gynaecol 1995;102(1):79-81.

Umbach DM, Wilcox AJ. A technique for measuring epidemiologically useful features of birthweight distributions. Stat Med 1996;15(13):1333-48.

Usher RH, Boyd ME, McLean FH, et al. Assessment of fetal risk in postdate pregnancies. Am J Obstet Gynecol 1988;158(2):259-64.

Valensise H, Dell'anna D, Menghini S, et al. Computerized analysis of fetal heart rate in postterm pregnancy. J Matern Fetal Invest 1994;4(4):251-6.

Van de Pas PM, Nijhuis JG, Jongsma HW. Fetal behaviour in uncomplicated pregnancies after 41 weeks of gestation. Early Hum. Dev. 1994;40(1):29-38.

Van GMJC, Major AL, Scherjon SA. Placental anatomy, fetal demise and therapeutic intervention in monochorionic twins and the transfusion syndrome: New hypotheses. Eur J Obstet Gynecol Reprod Biol 1998;78(1):53-62.

Vantanasiri C, Manassakorn J, Thitadilok W, et al. False reactive nonstress tests in postterm pregnancies. J Med Assoc Thai 1985;68(5):266-8.

Varaklis K, Gumina R, Stubblefield PG. Randomized controlled trial of vaginal misoprostol and intracervical prostaglandin E2 gel for induction of labor at term. Obstet Gynecol 1995;86(4 Pt 1):541-4.

Ventura SJ, Martin JA, Curtin SC, et al. Births: Final data for 1998. National Vital Statistics Report, vol 48, no 3. Hyattsville, MD. National Center for Health Statistics, 2000.

Villar J, Gulmezoglu AM, De OM. Nutritional and antimicrobial interventions to prevent preterm birth: An overview of randomized controlled trials. Obstet Gynecol Surv 1998;53(9):575-85.

Voss DH, Cumminsky KC, Cook VD, et al. Effects of three concentrations of intracervical prostaglandin Einf 2 gel for cervical ripening. J Matern Fetal Med 1996;5(4):186-93.

Votta RA, Cibils LA. Active management of prolonged pregnancy. Am J Obstet Gynecol 1993;168(2):557-63.

Wald NJ, Cuckle HS, Boreham J, et al. Maternal serum alpha-fetoprotein and birth weight. Br J Obstet Gynaecol 1980;87(10):860-3.

Waldenstrom U, Nilsson S, Fall O, et al. Effects of routine one-stage ultrasound screening in pregnancy: A randomised controlled trial. Lancet 1988;2(8611):585-8.

Walles B, Tyden T, Herbst A, et al. Maternal health care program and markers for late fetal death. Acta Obstet Gynecol Scand 1994;73(10):773-8.

Webb GW, Raynor BD, Huddleston JF, et al. Induction of labor with an unfavorable cervix: a randomized prospective trial. Am J Obstet Gynecol 1997;176:S22.

Webster LA, Daling JR, McFarlane C, et al. Prevalence and determinants of caesarean section in Jamaica. J Biosoc Sci 1992;24(4):515-25.

Weeks JW, Pitman T, Spinnato JA 2nd. Fetal macrosomia: does antenatal prediction affect delivery route and birth outcome? Am J Obstet Gynecol 1995;173(4):1215-9.

Weerasekera DS, Grant JM. Sweeping of the membranes is an effective method of induction of labour in prolonged pregnancy: a report of a randomised trial (2). Br J Obstet Gynaecol 1993;100(2):193-4.

Weiner Z, Farmakides G, Barnhard Y, et al. Doppler study of the fetal cardiac function in prolonged pregnancies. Obstet Gynecol 1996;88(2):200-2.

Weiner Z, Farmakides G, Schulman H, et al. Central and peripheral haemodynamic changes in post-term fetuses: Correlation with oligohydramnios and abnormal fetal heart rate pattern. Br J Obstet Gynaecol 1996;103(6):541-6.

Weiner Z, Farmakides G, Schulman H, et al. Computerized analysis of fetal heart rate variation in postterm pregnancy: prediction of intrapartum fetal distress and fetal acidosis. Am J Obstet Gynecol 1994;171(4):1132-8.

Weiner Z, Reichler A, Zlozover M, et al. The value of Doppler ultrasonography in prolonged pregnancies. Eur J Obstet Gynecol Reprod Biol 1993;48(2):93-7.

Weinstein D, Ezra Y, Picard R, et al. Expectant management of post-term patients: observations and outcome. J Matern Fetal Med 1996;5(5):293-7.

Weissman A, Yoffe N, Jakobi P, et al. Management of triplet pregnancies in the 1980s - Are we doing better? Am J Perinatol 1991;8(5):333-7.

Wikstrom I, Axelsson O, Bergstrom R, et al. Traumatic injury in large-for-date infants. Acta Obstet Gynecol Scand 1988;67(3):259-64.

Wilcox AJ, Weinberg CR, Baird DD. Timing of sexual intercourse in relation to ovulation. Effects on the probability of conception, survival of the pregnancy, and sex of the baby. N Engl J Med 1995;333(23):1517-21.

Wilcox JA, Nasrallah HA. Perinatal distress and prognosis of psychotic illness. Neuropsychobiology 1987;17(4):173-5.

Wing DA, Fassett MJ, Mishell DR. Mifepristone for preinduction cervical ripening beyond 41 weeks' gestation: a randomized controlled trial. Obstet Gynecol 2000;96(4):543-8.

Wing DA, Jones MM, Rahall A, et al. A comparison of misoprostol and prostaglandin E2 gel for preinduction cervical ripening and labor induction. Am J Obstet Gynecol 1995;172(6):1804-10.

Wing DA, Lovett K, Paul RH. Disruption of prior uterine incision following misoprostol for labor induction in women with previous cesarean delivery. Obstet Gynecol 1998;91(5 Pt 2):828-30.

Wing DA, Ortiz-Omphroy G, Paul RH. A comparison of intermittent vaginal administration of misoprostol with continuous dinoprostone for cervical ripening and labor induction. Am J Obstet Gynecol 1997;177(3):612-8.

Wing DA, Paul RH. A comparison of differing dosing regimens of vaginally administered misoprostol for preinduction cervical ripening and labor induction [published erratum appears in Am J Obstet Gynecol 1997 Jun;176(6):1423]. Am J Obstet Gynecol 1996;175(1):158-64.

Wing DA, Rahall A, Jones MM, et al. Misoprostol: an effective agent for cervical ripening and labor induction. Am J Obstet Gynecol 1995;172(6):1811-6.

Wiriyasirivaj B, Vutyavanich T, Ruangsri RA. A randomized controlled trial of membrane stripping at term to promote labor. Obstet Gynecol 1996;87(5 Pt 1):767-70.

Wischnik A, Zieger W. Tocolytic therapies in development for preterm labour. Expert Opin Invest Drugs 1994;3(6):621-9.

Witter FR, Weitz CM. A randomized trial of induction at 42 weeks gestation versus expectant

management for postdates pregnancies. Am J Perinatol 1987;4(3):206-11.

Witter FR, Weitz CM. Cervical examination prior to induction in postdate pregnancies. Surg Gynecol Obstet 1989;168(3):214-6.

Wood CL. Postdate pregnancy update. [Review]. J Nurse Midwifery 1994;39(2 Suppl):110S-22S.

Wood L. Induction of labour confers benefit in prolonged pregnancy. Br J Obstet Gynaecol 1994;101(12):1102.

Wu MY, Yang YS, Huang SC, et al. Prolongation of pregnancy and survival of both twins after preterm premature rupture of membrane of twin A and 19 weeks' gestation. Acta Obstet Gynecol Scand 1996;75(3):299-302.

Xenakis EM, Piper JM, Conway DL, et al. Induction of labor in the nineties: conquering the unfavorable cervix. Obstet Gynecol 1997;90(2):235-9.

Yates WT, Cibils LA. Why not begin 'active management' of prolonged pregnancy sooner? (2). Am J Obstet Gynecol 1994;170(1 I):253.

Yeast JD, Jones A, Poskin M. Induction of labor and the relationship to cesarean delivery: A review of 7001 consecutive inductions. Am J Obstet Gynecol 1999;180(3 Pt 1):628-33.

Yeh S, Huang X, Phelan JP. Postterm pregnancy after previous cesarean section. J Reprod Med 1984;29(1):41-4.

Yeh SY, Bruce SL, Thornton YS. Intrapartum monitoring and management of the postdate fetus. [Review]. Clin Perinatol 1982;9(2):381-6.

Yeh SY, Read JA. Management of post-term pregnancy in a large obstetric population. Obstet Gynecol 1982;60(3):282-7.

Yeh SY, Read JA. Plasma unconjugated estriol as an indicator of fetal dysmaturity in postterm pregnancy. Obstet Gynecol 1983;62(1):22-5.

Yoonessi M, Wieckowska W, Mariniello D, et al. Cervical intra-epithelial neoplasia in pregnancy. Int J Gynecol Obstet 1982;20(2):111-8.

Yudkin P. Risk of unexplained stillbirth in prolonged pregnancy. Midwife Health Visit Community Nurse 1988;24(10):407-10.

Yudkin PL, Redman CW. Caesarean section dissected, 1978-1983. Br J Obstet Gynaecol 1986;93(2):135-44.

Yudkin PL, Wood L, Redman CWG. Risk of unexplained stillbirth at different gestational ages. Lancet 1987;1:1192-4.

Zelop CM, Shipp TD, Cohen A, et al. Trial of labor after 40 weeks' gestation in women with prior cesarean. Obstet Gynecol 2001;97:391-3.

Zhao YG, Shao YL. The combined tests of overnight 12-hour urine E3/C ratio and fetal heart rate in perinatal monitoring. Chin Med J (Engl) 1989;102(3):188-92.

Zhou W, Sorensen HT, Olsen J. Induced abortion and subsequent pregnancy duration. Obstet Gynecol 1999;94(6):948-53.

Zuazu J, Julia A, Sierra J, et al. Pregnancy outcome in hematologic malignancies. Cancer 1991;67(3):703-9.

Appendix 1: Data-Abstraction Form

POST-TERM PREGNANCY
ARTICLE ABSTRACTING FORM

Reviewer:_____First Author:_____Year:_____Procite #:_____

ARTICLE FOCUS (circle one): Testing / Management / Both

STUDY DESIGN (check one):

_____RCT – Randomization method: _____Sealed envelope

_____Date/Chart #

_____Not described

_____Other – describe:_____

_____Cohort

_____Case series, no controls, n = _____

_____Case series, historical controls, n = _____

_____Case series, concomitant controls, n = _____

_____Not specified or unable to classify

REASSESSMENT:

Recode article as:_____ Exclude (give reason):_____

*Note: All non-RCTs should be **excluded** from the management review*

KEY QUESTIONS ADDRESSED (check all that apply):

_____1. What are the test characteristics (reliability, sensitivity, specificity, predictive values) and costs of measures used in the management of postdates pregnancy: (a) to assess risks to the fetus of postdates pregnancy, and (b) to assess the likelihood of a successful induction?

_____2. What are the benefits, risks, and costs of currently available interventions for induction of labor?

_____3. What is the direct evidence comparing the benefits, risks, and costs of planned induction versus expectant management at various gestational ages?

_____4. Are the epidemiology and outcomes of postdates pregnancy different for women in different ethnic groups, different socioeconomic groups, or in adolescent women?

STUDY LOGISTICS:

Inclusive dates of data collection (give month and year): from_____to_____

Multicenter study? (circle one): Yes / No If "Yes," no. of sites:_____

Geographic location (in US, give city and state; outside of US, give city and country. If multicenter trial or network, give name, e.g., NICHD MFM Network, RADIUS):_____

TYPES OF PROVIDERS (check all that apply):	**STUDY SETTING (check all that apply):**
_____Unspecified OB/GYN	_____University hospital
_____General OB/GYN	_____Community hospital
_____MFM	_____Unspecified hospital
_____Family practice	_____Freestanding birthing center
_____Nurse midwives	_____Outpatient clinic/physician office
_____Other midwives	_____Not specified or unable to determine
_____Other – describe:_____	_____Other – describe:_____
_____Not specified	

GESTATIONAL AGE DETERMINED BY (check all that apply):

_____LMP

_____1st trimester U/S

_____2nd trimester U/S

_____Other – specify: _____

INCLUSION CRITERIA:	**EXCLUSION CRITERIA:**

SUBJECT CHARACTERISTICS:

1) Identify interventions A, B, and C, and indicate which (if any) served as control

2) Use "NR" to indicate "Not reported"

	Intervention A =	Intervention B =	Intervention C =	Overall
AGE (specify summary statistic [mean, median] and measure of dispersion [standard deviation, range, etc.]; if age not described in these terms, then enter as reported):				
Mean:				
Median:				
SD:				
Range:				
RACE (specify distribution):				
White:	n = / %	n = / %	n = / %	n = / %
Black:	n = / %	n = / %	n = / %	n = / %
Hispanic:	n = / %	n = / %	n = / %	n = / %
Other:	n = / %	n = / %	n = / %	n = / %
GESTATIONAL AGE AT ENTRY INTO STUDY (specify either summary statistic [mean, median] and measure of dispersion [SD, range] or percent in each category; indicate whether measured in days or weeks)				
PARITY (specify either summary statistic [mean, median] and measure of dispersion [SD, range] or percentage in each category):				
BISHOP SCORE (specify either summary statistic [mean, median] and measure of dispersion [SD, range] percentage in each category):				
OTHER measure of cervical dilatation or effacement (specify):				

INTERVENTIONS

Describe the testing and management interventions used in each study group. Include all information necessary to reproduce the treatment/monitoring/testing algorithms used. For example:

Sample Intervention A = Induction

If cervix < 3 cm dilated and < 50% effaced and fetal heart rate normal, then pt given PGE2 gel (Prepidil) 0.5 mg intracervically – max of 3 doses at 6-hr intervals – fetus monitored continuously for min of 1 hr after insertion of gel

If gel not used or did not induce labor within 12 hrs of insertion of last dose, then labor induced by IV oxytocin or amniotomy or both

Interventions to be considered include:

1) **Tests of fetal well-being: No tests, nonstress test, biophysical profile, contraction stress test, amniotic fluid volume, uterine vessel Doppler flow, other, combinations of the preceding**
2) **Tests of fetal size: Physical exam, ultrasound, other**
3) **Tests of readiness for delivery: Bishop score, fetal fibronectin, other, combinations of the preceding**
4) **Interventions: Monitoring/conservative care, stripping of membranes, oxytocin, prostaglandin gel, misoprostil, mechanical interventions**

Intervention A =

Intervention B =

Intervention C =

PATIENT NUMBERS, DROPOUTS AND LOSS TO FOLLOW-UP:

Outcome	Intervention A =	Intervention B =	Intervention C =
No. of subjects at start:			
No. of subjects who did not receive allocated intervention due to:			
Spontaneous labor:	n = / %	n = / %	n = / %
Other complications:	n = / %	n = / %	n = / %
Other/unspecified causes:	n = / %	n = / %	n = / %
No. of subjects at end who had received allocated intervention:	n = / %	n = / %	n = / %
Any post-discharge follow-up? (circle one)	Yes / No	Yes / No	Yes / No
No. of subjects lost to post-discharge follow-up:	n = / %	n = / %	n = / %

MANAGEMENT OUTCOMES:

Outcome Measured (Describe)	How measured, (e.g., scale/units used, %)	Intervention A =	Intervention B =	Intervention C =	P value
FETAL OUTCOMES (e.g., stillbirth, Apgar scores, admission to NICU, shoulder dystocia, weight, etc.):					
1)					
2)					
3)					
4)					
5)					

MANAGEMENT OUTCOMES (continued):

Outcome Measured (Describe)	How measured, (e.g., scale/units used, %)	Intervention A =	Intervention B =	Intervention C =	P value
FETAL OUTCOMES (continued)					
6)					
7)					
MATERNAL OUTCOMES (e.g., maternal trauma, C-section rate [with causes], infection, etc.):					
1)					
2)					
3)					
4)					
5)					
6)					
7)					
OTHER OUTCOMES					
1)					
2)					

TEST PERFORMANCE OUTCOMES (Testing Articles Only):

Comparison 1

	Reference standard/outcome =			
Screening test =	Ref standard result 1 =	Ref standard result 2 =	Ref standard result 3 =	Totals:
Screen test result 1 =				
Screen test result 2 =				
Screen test result 3 =				
Totals:				

Comparison 2

	Reference standard/outcome =			
Screening test =	Ref standard result 1 =	Ref standard result 2 =	Ref standard result 3 =	Totals:
Screen test result 1 =				
Screen test result 2 =				
Screen test result 3 =				
Totals:				

Comparison 3

	Reference standard/outcome =			
Screening test =	Ref standard result 1 =	Ref standard result 2 =	Ref standard result 3 =	Totals:
Screen test result 1 =				
Screen test result 2 =				
Screen test result 3 =				
Totals:				

Other test performance results (including sensitivity and specificity and qualitative results):

COST/CHARGES/RESOURCE UTILIZATION OUTCOMES:

Outcome Measured	How measured, (e.g., scale/units used, %)	Intervention A =	Intervention B =	Intervention C =	P value
Total costs/intervention:					
Mean:					
Median:					
SD:					
Range:					
Other cost/resource outcome (specify):					

QUALITY SCORE:
(Check "Yes" or "No" for each item)

Type of Article	Yes	No
MANAGEMENT ARTICLES		
Randomized assignment to intervention?		
Randomization method clearly described and appropriate?		
Study population similar to likely patient population?		
Intervention protocols clearly described or referenced?		
Description provided of how decisions made about mode of delivery?		
Statistical issues addressed/discussed:		
Sample size?		
Use of appropriate tests?		
Study population characterized by:		
Gestational age?		
Dating criteria specified?		
Bishop score or other measure of cervical ripeness?		
TESTING ARTICLES		
Reference standard defined?		
Randomized assignment to test?		
Randomization method clearly described and appropriate?		
Verification bias assessed or discussed?		
Test reliability/variability addressed or discussed?		
Study population well characterized by:		
Gestational age?		
Dating criteria specified?		
Absence of other risk factors (diabetes, HTN, etc.)?		
Study population similar to likely patient population?		
Testing protocol clearly described or referenced?		
Statistical issues addressed/discussed:		
Sample size?		
Use of appropriate tests?		

Appendix 2: Evidence Table Templates

Template for Evidence Table 1

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Author and Pro-Cite #	Design: [RCT, etc., including description of method of randomization] Test(s) studied: 1) 2) 3) etc. Reference standard(s): 1) 2) etc. Dates: Location: Setting: [including whether single- or multicenter] Type(s) of providers: Length of follow-up:	No. of subjects at start: Dropouts: Loss to follow-up: No. of subjects at end: Inclusion criteria: Exclusion criteria: Age: Race: Gestational age at entry: Dating criteria: Parity: Bishop score: Other: [including other measures of cervical ripeness]	1) 2) 3) 4) 5) 6) 7) 8) 9) 10) 11) 12) 13) 14) 15)	1) Outcome1: 2) Outcome2: 3) Outcome3: 4) Outcome4: 5) Outcome5: 6) Outcome6: 7) Outcome7: 8) Outcome8: 9) Outcome9: 10) Outcome10: 11) Outcome11: 12) Outcome12: 13) Outcome13: 14) Outcome14: 15) Outcome15:	QUALITY SCORES: TESTING Reference standard: Randomized: Method of randomization: Verification bias: Test reliability/variability: Gestational age: Dating criteria: Other risk factors absent: Similar to likely pt pop: Testing protocol described: Sample size: Statistical tests: MANAGEMENT Randomized: Method of randomization: Similar to likely pt pop: Interventions described: Mode of delivery: Sample size: Statistical tests: Gestational age: Dating criteria: Bishop score:

Template for Evidence Tables 2 and 3

Study	Design and Interventions	Patient Population	Outcomes Reported	Results	Quality Score/Notes
Author and Pro-Cite #	Design: [RCT, etc., including description of method of randomization] Interventions: 1) 2) 3) etc. Dates: Location: Setting: [including whether single- or multicenter] Type(s) of providers: Length of follow-up:	No. of subjects at start: Dropouts: Loss to follow-up: No. of subjects at end: Inclusion criteria: Exclusion criteria: Age: Race: Gestational age at entry: Dating criteria: Parity: Bishop score: Other: [including other measures of cervical ripeness]	1) 2) 3) 4) 5) 6) 7) 8) 9) 10) 11) 12) 13) 14) 15)	1) Outcome1: 2) Outcome2: 3) Outcome3: 4) Outcome4: 5) Outcome5: 6) Outcome6: 7) Outcome7: 8) Outcome8: 9) Outcome9: 10) Outcome10: 11) Outcome11: 12) Outcome12: 13) Outcome13: 14) Outcome14: 15) Outcome15:	QUALITY SCORE: Randomized: Method of randomization: Similar to likely pt pop: Interventions described: Mode of delivery: Sample size: Statistical tests: Gestational age: Dating criteria: Bishop score: