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.

4

- 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.

AHRQ

www.ahrq.gov