Evidence-based Practice Center Systematic Review Protocol

Project Title: *Cervical Ripening in the Outpatient Setting*

I. Background

Induction of labor (IOL) is the process of initiating labor by using medications, mechanical (devices), or other methods, with a goal to achieve safe vaginal birth.[1] IOL has shown maternal/child benefit when the health of a pregnant woman or fetus is at risk (e.g. maternal hypertension or diabetes, fetal growth restriction, and in postterm pregnancies).[2,3] In addition to these medically-indicated deliveries, IOL is increasingly being done on an elective basis, primarily for reasons of scheduling at the request of pregnant women, or to insure availability of appropriate providers.[4] A recent large randomized study of low-risk nulliparous pregnant women (the ARRIVE trial)[5] demonstrated that induction of labor at 39 weeks, compared to expectant management, resulted in lower cesarean delivery (CD) rates and no difference in neonatal outcomes.[2,3] IOL rates are rising dramatically in the United States reaching 25.7 percent in 2017[6] , with tertiary care centers that provide high-risk obstetric care, such as OHSU, reaching 46.5 percent in 2019.[7] Labor induction occurs in approximately one-quarter of term pregnancies, with estimates of 77 to 85 percent occurring due to medical indications.[8-10]

Cervical ripening (CR), often an initial component of labor induction, is the process of softening and effacing the cervix as well as stimulating early cervical dilation. Based on data from trials of labor induction, approximately 83 to 85 percent of women with an indication for induction require cervical ripening.[11,12] Common CR methods include pharmacologic options, such as prostaglandins (misoprostol and Prostaglandin E2), and mechanical options, such as inflating a balloon catheter in the cervix (e.g., Foley or Cook catheter).

Traditionally CR has been performed as an inpatient procedure, and while there is variation, it can require substantial time and resources to accomplish successfully, due to multiple factors. While prostaglandins (vaginal or oral) and mechanical methods (e.g. balloon catheters) are the most commonly used methods of CR in the inpatient setting, there is variation in the dose, regimen, or protocols applied. Some women's cervixes will rapidly respond to a CR intervention, while others require extended time with more than one intervention being tried if the first one fails. While interventions used for CR are generally not costly, the hospital inpatient resources used, including highly skilled labor and delivery staff, contribute to increased costs when CR care is provided in the inpatient setting.

Because of the time involved, many women would prefer to be at home during the CR process, and because of the resources and variation involved, providers are also interested in exploring safe methods of CR in the outpatient setting. Informed by these considerations, there is growing interest in and evidence for outpatient CR. It has been proposed that outpatient CR may facilitate more efficient and more satisfying IOL, while reducing inpatient length of stay compared to inpatient CR.

The risks of CR are similar to those of spontaneous labor, compounded by known and theoretical iatrogenic effects of medication and mechanical cervical stimulation. However, there are concerns regarding potential risks of outpatient CR to the woman and fetus in comparison with CR in the inpatient setting. The risks of CR may be mitigated through the choice of CR method and clinical management. For example, fetal monitoring is recommended with prostaglandins because use of these medications have been associated with risk of tachysystole and fetal distress. Careful review of existing literature is needed to elucidate whether these risks occur more frequently when CR is accomplished in the outpatient versus inpatient setting and whether maternal or fetal characteristics differentially affect these risks. In addition, understanding the range of feasible and safe outpatient CR options, and what form of fetal surveillance should be used (if any), is an important aspect of this review.

A woman's preferences and satisfaction related to the setting of CR also need to be considered. Some may actively seek outpatient CR and others may strongly prefer inpatient CR. This likely variation in preferences and satisfaction has been identified as an important contextual question of this review.

Despite potential cost saving and sometimes strong personal preferences favoring outpatient CR, this approach to care is still debated. Controversy is driven by interpretation of risk, clinician's discipline and experience (e.g., obstetrician vs. midwife),[13,14] and geographical practice variation. Clinician and institutional risk-aversion driven by potential legal litigation is also a consideration. The 2009 American College of Obstetricians and Gynecologists (ACOG) Practice Bulletin on induction of labor summarized evidence on CR in the outpatient setting, based on only two studies available at that time (one on a prostaglandin, one on a balloon catheter),[15] ultimately not reaching a recommendation. A 2017 Cochrane review found that evidence on outpatient vs inpatient CR was insufficient to address differences in maternal and fetal health outcomes, such as CD, between settings.[16] This review included only randomized controlled trials, and included interventions not available in the US or that are used primarily to stimulate or maintain contractions rather than primarily for CR. Many CR studies have been conducted in non-U.S. settings, where patient acceptance and understanding of risk may be different, in addition to variation in provider philosophy and health system resources. There is a need to assess the benefits of outpatient vs. inpatient CR, without increasing risk (rise in CD rate, adverse neonatal outcomes), framed within considerations of cost, patient autotomy, and satisfaction. This is the crux of the decisional dilemma, when CR is indicated what methods can be recommended as safe and effective in the outpatient setting and what surveillance best serves women induced with prostaglandin in the outpatient setting.

We anticipate that the evidence on outpatient CR presents some specific challenges, including (1) the characteristics of patients in studies of outpatient CR may be limited to low-risk pregnancies, (2) there is significant heterogeneity in protocols used for CR regardless of setting, (3) adverse maternal and neonatal outcomes are rare and studies may not have adequate statistical power to detect differences, (4) some commonly reported outcomes have varying definitions, e.g. time to delivery, (5) the clinical meaningfulness of intermediate or surrogate outcomes are unclear, and (6) women have strong birth experience preferences that vary geographically, ethnically and culturally—patient satisfaction in one patient population may not be generalizable. These factors may limit the ability to combine studies in meaningful ways, and the applicability to other patients or settings.

Purpose of the Review

This systematic review will assess the comparative effectiveness and potential harms of cervical ripening in the outpatient versus the inpatient setting. The intended audience includes the ACOG's guideline developers, family physicians, practitioners who deliver infants (e.g. obstetricians, nurse midwives), other personnel who administer and monitor cervical ripening and health system policymakers. In addition to these clinical implications, we hope to inform the future research necessary to provide high-quality, evidence-based care to all pregnant women.

II. Key Questions

An initial set of Key Questions (KQs) were posted on the Agency for Healthcare Research and Quality (AHRQ) Effective Health Care Program website for public input from May 10 to May 30, 2019, prior to the initiation of this review, and a public stakeholder webinar was held by the Patient Centered Outcomes Research Institute (PCORI). Changes to the Key Questions based on public comment include expanding defined subgroups and removing specific brand names for mechanical devices. Subsequently, a group of Key Informants and a separate group of Technical Experts, including representatives of ACOG's guideline group, provided comments on the scope of the review. The following Key Questions and inclusion criteria reflect these suggestions.

KQ1: How do the effectiveness and harms of CR using prostaglandins compare in the outpatient vs. inpatient setting?

1a: How do effectiveness and harms vary by choice of prostaglandin?

1b: Do effectiveness and harms vary by important patient characteristics (such as gestational age, parity, uncomplicated pregnancy, prior cesarean delivery, etc.)?

KQ2: How do the effectiveness and harms of CR using mechanical methods (e.g., balloon catheters) compare in the outpatient vs. inpatient setting?

2a: How do effectiveness and harms vary by choice of mechanical method in the inpatient versus the outpatient setting?

2b: Do effectiveness and harms vary by important patient characteristics (such as gestational age, parity, uncomplicated pregnancy, prior cesarean delivery, etc.)?

KQ3: How do the effectiveness and harms of CR in the *outpatient setting* vary by method of CR compared with each other?

3a: Do effectiveness and harms vary by important patient characteristics (such as gestational age, parity, uncomplicated pregnancy, prior cesarean delivery, etc.)?

KQ4: How do the effectiveness and harms of different methods and protocols for fetal surveillance compare with each other or with no monitoring in pregnant women undergoing CR with prostaglandins?

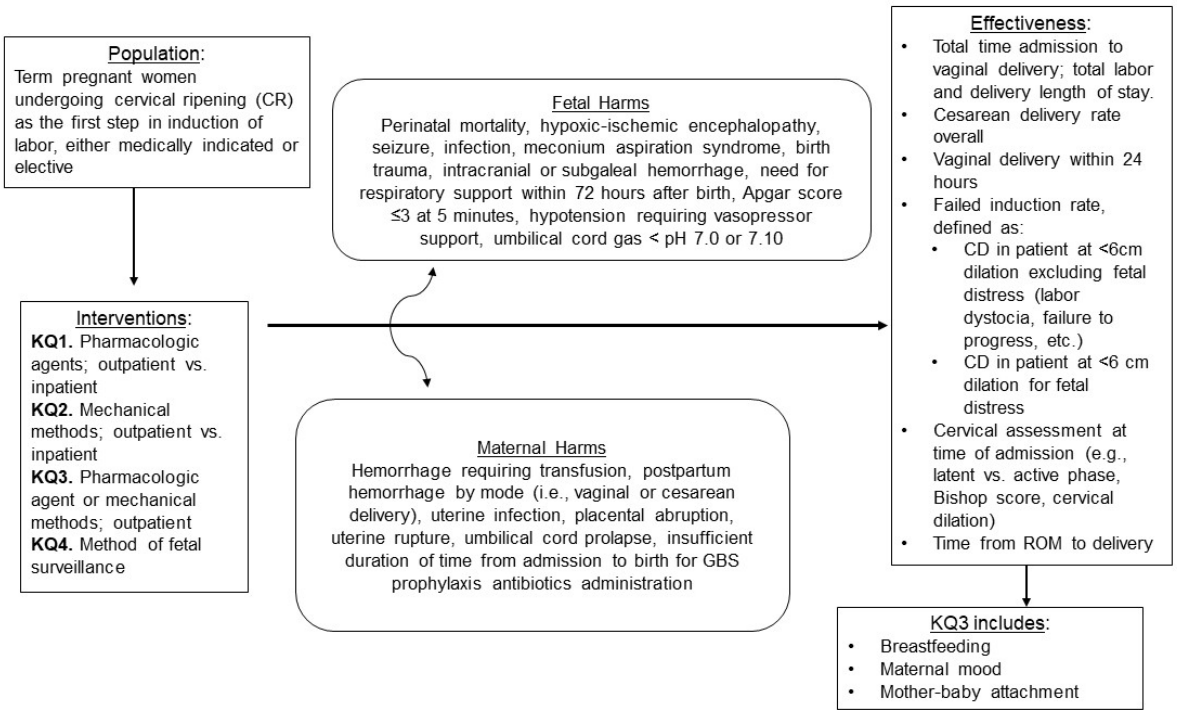
4a. Do effectiveness and harms vary by important patient characteristics (such as gestational age, parity, uncomplicated pregnancy, prior cesarean delivery, etc.)?

Contextual Question: What evidence informs preference for or tolerability of different methods of CR in the outpatient setting or outpatient compared to the inpatient setting?

Following the methods of the US Preventive Services Task Force (USPSTF)[17], a contextual question represents issues in a review for which a valid, but not necessarily systematic, summary of current research is needed in order to provide context on the issue. See the Methods section below for more details.

III. Analytic Framework

Figure 1. Analytic Framework



CR = cervical ripening; CD = Cesarean delivery; ROM = rupture of membrane; KQ = Key Question

IV. Methods

All methods used for this systematic review will be conducting in accordance with the Agency for Healthcare Research and Quality's *Methods Guide for Effectiveness and Comparative Effectiveness Review*,[18] developed for the Evidence-based Practice Centers.

Criteria for Inclusion/Exclusion of Studies in the Review

The criteria for inclusion and exclusion of studies for the systematic review will be based on the Key Questions and on the specific criteria listed in Table 1. Population, interventions, comparators, outcomes, timing, and settings (PICOTS) for each key question. Outcomes prioritized as primary outcomes for this systematic review are footnoted and listed in **bold** below.

Table 1. PICOTS

PICOTS	Inclusion Key Question 1: Prostaglandin Inpatient vs. Outpatient	Inclusion Key Question 2: Mechanical Method Inpatient vs Outpatient	Inclusion Key Question 3: Outpatient comparison of methods	Inclusion Key Question 4: Fetal Surveillance	Exclusion
Population	• Pregnant women ≥37 weeks undergoing CR in the outpatient setting • Important maternal subgroups: parity, maternal age, GBS status, diabetes (pre-gestational, gestational), hypertension (chronic, preeclampsia without severe features, gestational) • Important fetal subgroups: fetal growth restriction, gestational age (<39 weeks, 39 to 41 weeks, >41 weeks)	• Pregnant women ≥37 weeks undergoing CR in the outpatient setting • Important maternal subgroups: parity, maternal age, GBS status, diabetes (pre-gestational, gestational), hypertension (chronic, preeclampsia without severe features, gestational) • Important fetal subgroups: fetal growth restriction, gestational age (<39 weeks, 39 to 41 weeks, >41 weeks)	• Pregnant women ≥37 weeks undergoing CR in the outpatient setting • Important maternal subgroups: parity, maternal age, GBS status, diabetes (pre-gestational, gestational), hypertension (chronic, preeclampsia without severe features, gestational) • Important fetal subgroups: fetal growth restriction, gestational age (<39 weeks, 39 to 41 weeks, >41 weeks)	• Pregnant women ≥37 weeks undergoing CR with a prostaglandin • Important maternal subgroups: parity, maternal age, GBS status, diabetes (pre-gestational, gestational), hypertension (chronic, preeclampsia without severe features, gestational) • Important fetal subgroups: fetal growth restriction, gestational age (<39 weeks, 39 to 41 weeks, >41 weeks)	Women with contraindications to CR in the outpatient setting: a multiple pregnancy, prior uterine rupture and breech presentation of the fetus.
Intervention	• Pharmacologic agents (prostaglandins) given in outpatient setting	• Mechanical methods (balloon catheters, laminaria tents) used in outpatient setting	Mechanical methods (balloon catheters, laminaria tents) or pharmacologic agents (prostaglandins)	• Any method of fetal surveillance	• Catheters not available in the U.S. • Pharmacy-compounded prostaglandin products • Other CR methods: Castor oil, nipple stimulation, membrane stripping, sexual intercourse, acupuncture/pressure, transcutaneous nerve stimulation, herbal compounds

PICOTS	Inclusion Key Question 1: Prostaglandin Inpatient vs. Outpatient	Inclusion Key Question 2: Mechanical Method Inpatient vs Outpatient	Inclusion Key Question 3: Outpatient comparison of methods	Inclusion Key Question 4: Fetal Surveillance	Exclusion
Comparator	• Mechanical (i.e., balloon catheters, luminaria tents) and/or pharmacologic (i.e., prostaglandins) methods in the inpatient setting	• Mechanical (i.e., balloon catheters, luminaria tents) and/or pharmacologic (i.e., prostaglandins) methods in the inpatient setting	• Any comparator including alternative mechanical device or protocol, alternative pharmacologic agent or dose, combination mechanical and pharmacologic, placebo, and other CR methods excluded as intervention (e.g., Castor oil, acupuncture)	• Another method of fetal surveillance • Another protocol for fetal surveillance with the same method • No monitoring	• Catheters not available in the U.S. • Pharmacy-compounded prostaglandin products

PICOTS	Inclusion Key Question 1: Prostaglandin Inpatient vs. Outpatient	Inclusion Key Question 2: Mechanical Method Inpatient vs Outpatient	Inclusion Key Question 3: Outpatient comparison of methods	Inclusion Key Question 4: Fetal Surveillance	Exclusion
Outcomes Effectiveness (birth-related)	• **Total time admission to vaginal delivery; total L&D length of stay[c]** • **Cesarean delivery rate overall[c]** • Vaginal delivery within 24 hours • Failed induction rate, defined as: o CD in patient at <6cm dilation excluding fetal distress (labor dystocia, failure to progress, etc.) o CD in patient at <6 cm dilation for fetal distress • Cervical assessment at time of admission (e.g., latent vs. active phase, Bishop score, cervical dilation) • Time from ROM to delivery	• **Total time admission to vaginal delivery; total L&D length of stay[c]** • **Cesarean delivery rate overall[c]** • Vaginal delivery within 24 hours • Failed induction rate, defined as: o CD in patient at <6cm dilation excluding fetal distress (labor dystocia, failure to progress, etc.) o CD in patient at <6 cm dilation for fetal distress • Cervical assessment at time of admission (e.g., latent vs. active phase, Bishop score, cervical dilation) • Time from ROM to delivery	• **Total time admission to vaginal delivery; total L&D length of stay[c]** • **Cesarean delivery rate overall[c]** • Vaginal delivery within 24 hours • Failed induction rate, defined as: o CD in patient at <6cm dilation excluding fetal distress (labor dystocia, failure to progress, etc.) o CD in patient at <6 cm dilation for fetal distress • Cervical assessment at time of admission (e.g., latent vs. active phase, Bishop score, cervical dilation) • Time from ROM to delivery • Breastfeeding[b] • Maternal mood[b] • Mother-baby attachment[b]	• **Total time admission to vaginal delivery; total L&D length of stay[c]** • **Cesarean delivery rate overall[c]** • Vaginal delivery within 24 hours • Failed induction rate, defined as: o CD in patient at <6cm dilation excluding fetal distress (labor dystocia, failure to progress, etc.) o CD in patient at <6 cm dilation for fetal distress • Cervical assessment at time of admission (e.g., latent vs. active phase, Bishop score, cervical dilation) • Time from ROM to delivery	Outcomes not listed in inclusion criteria

PICOTS	Inclusion Key Question 1: Prostaglandin Inpatient vs. Outpatient	Inclusion Key Question 2: Mechanical Method Inpatient vs Outpatient	Inclusion Key Question 3: Outpatient comparison of methods	Inclusion Key Question 4: Fetal Surveillance	Exclusion
Outcomes Fetal Harms	• **Perinatal Mortality**[c] • **Hypoxic-ischemic**[c] **encephalopathy**[c] • **Seizure**[c] • **Infection** (confirmed sepsis or pneumonia)[c] • **Meconium aspiration syndrome**[c] • **Birth trauma** (e.g., bone fracture, neurologic injury, or retinal hemorrhage)[c] • **Intracranial or subgaleal hemorrhage**[c] • Need for respiratory support within 72 hours after birth • Apgar score ≤3 at 5 minutes[a] • Hypotension requiring vasopressor support • Umbilical cord gas < pH 7.0 or 7.10	• **Perinatal Mortality**[c] • **Hypoxic-ischemic encephalopathy**[c] • **Seizure**[c] • **Infection** (confirmed sepsis or pneumonia)[c] • **Meconium aspiration syndrome**[c] • **Birth trauma** (e.g., bone fracture, neurologic injury, or retinal hemorrhage)[c] • **Intracranial or subgaleal hemorrhage**[c] • Need for respiratory support within 72 hours after birth • Apgar score ≤3 at 5 minutes[a] • Hypotension requiring vasopressor support • Umbilical cord gas < pH 7.0 or 7.10	• **Perinatal Mortality**[c] • **Hypoxic-ischemic encephalopathy**[c] • **Seizure**[c] • **Infection** (confirmed sepsis or pneumonia)[c] • **Meconium aspiration syndrome**[c] • **Birth trauma** (e.g., bone fracture, neurologic injury, or retinal hemorrhage)[c] • **Intracranial or subgaleal hemorrhage**[c] • Need for respiratory support within 72 hours after birth • Apgar score ≤3 at 5 minutes[a] • Hypotension requiring vasopressor support • Umbilical cord gas < pH 7.0 or 7.10	• **Perinatal Mortality**[c] • **Hypoxic-ischemic encephalopathy**[c] • **Seizure**[c] • **Infection** (confirmed sepsis or pneumonia)[c] • **Meconium aspiration syndrome**[c] • **Birth trauma** (e.g., bone fracture, neurologic injury, or retinal hemorrhage)[c] • **Intracranial or subgaleal hemorrhage**[c] • Need for respiratory support within 72 hours after birth • Apgar score ≤3 at 5 minutes[a] • Hypotension requiring vasopressor support • Umbilical cord gas < pH 7.0 or 7.10	Outcomes not listed in inclusion criteria

PICOTS	Inclusion Key Question 1: Prostaglandin Inpatient vs. Outpatient	Inclusion Key Question 2: Mechanical Method Inpatient vs Outpatient	Inclusion Key Question 3: Outpatient comparison of methods	Inclusion Key Question 4: Fetal Surveillance	Exclusion
Outcomes Maternal Harms	• **Hemorrhage requiring transfusion**[c] • **Postpartum hemorrhage by mode** (vaginal, cesarean)[c] • **Uterine infection** (i.e., choriamnionitis, administration of antibiotics in labor other than GBS prophylaxis)[c] • Placental abruption Uterine rupture • Umbilical cord prolapse • Duration of time between hospital admission to birth that is insufficient to enable complete GBS prophylaxis antibiotics administration per CDC guidelines	• **Hemorrhage requiring transfusion**[c] • **Postpartum hemorrhage by mode** (vaginal, cesarean)[c] • **Uterine infection** (i.e., choriamnionitis, administration of antibiotics in labor other than GBS prophylaxis)[c] • Placental abruption • Uterine rupture • Umbilical cord prolapse • Duration of time between hospital admission to birth that is insufficient to enable complete GBS prophylaxis antibiotics administration per CDC guidelines	• **Hemorrhage requiring transfusion**[c] • **Postpartum hemorrhage by mode** (vaginal, cesarean)[c] • **Uterine infection** (i.e., choriamnionitis, administration of antibiotics in labor other than GBS prophylaxis)[c] • Placental abruption Uterine rupture • Umbilical cord prolapse • Duration of time between hospital admission to birth that is insufficient to enable complete GBS prophylaxis antibiotics administration per CDC guidelines	• **Hemorrhage requiring transfusion**[c] • **Postpartum hemorrhage by mode** (vaginal, cesarean)[c] • **Uterine infection** (i.e., choriamnionitis, administration of antibiotics in labor other than GBS prophylaxis)[c] • Placental abruption • Uterine rupture • Umbilical cord prolapse • Duration of time between hospital admission to birth that is insufficient to enable complete GBS prophylaxis antibiotics administration per CDC guidelines	Outcomes not listed in inclusion criteria
Timing	Maternal outcomes • From CR initiation to within 1-week following delivery Infant outcomes • Immediately following delivery	Maternal outcomes • From CR initiation to within 1-week following delivery Infant outcomes • Immediately following delivery.	Maternal and additional outcomes (i.e., breastfeeding, maternal mood, mother-baby attachment) • From CR initiation to 1-year postpartum Infant outcomes • Immediately following delivery	Maternal outcomes • From CR initiation to within 1-week following delivery Infant outcomes • Immediately following delivery	KQ 1,2,4: Outcomes occurring after 1-week post delivery KQ3: Outcomes for breastfeeding, mother-infant attachment, and maternal mood occurring after 1 year post-delivery.
Setting	• Inpatient versus outpatient settings	• Inpatient versus outpatient settings	• Outpatient setting	• Inpatient and outpatient settings	

PICOTS	Inclusion Key Question 1: Prostaglandin Inpatient vs. Outpatient	Inclusion Key Question 2: Mechanical Method Inpatient vs Outpatient	Inclusion Key Question 3: Outpatient comparison of methods	Inclusion Key Question 4: Fetal Surveillance	Exclusion
Study design	• Randomized Controlled Trials; recent high quality Systematic Reviews; if RCT evidence for benefits is insufficient, include large, high quality cohort studies comparing inpatient and outpatient setting. • Include high quality cohort and case-control studies for harms.	• Randomized Controlled Trials; recent high quality Systematic Reviews; if RCT evidence for benefits is insufficient, include large, high quality cohort studies comparing inpatient and outpatient setting. • Include high quality cohort and case-control studies for harms.	• Randomized Controlled Trials; recent high quality Systematic Reviews; if RCT evidence for benefits is insufficient, include large, high quality cohort studies comparing inpatient and outpatient setting. • Include high quality cohort and case-control studies for harms.	• Randomized Controlled Trials; recent high quality Systematic Reviews; if RCT evidence for benefits is insufficient, include large, high quality cohort studies comparing inpatient and outpatient setting. • Include high quality cohort and case-control studies for harms.	Case series, pre-post studies, case reports

[a] Will Consider higher thresholds from older studies if inadequate evidence with this threshold
[b] Specific to Key Question 3
[c] (Bolded) items indicate Primary Outcomes
CR = cervical ripening; CD = cesarean delivery; KQ = Key Question; ROM = rupture of membrane; CDC = Centers for Disease Control and Prevention; L&D = labor and delivery; RCTs = randomized controlled trials

Study Design: For all Key Questions, we will include randomized controlled trials (RCTs) for benefits and harms, and additionally large (N>200) comparative cohort or case-control studies to further evaluate harms. If evidence on benefits from RCTs is inconclusive for a key question or subquestion, comparative observational studies may be considered with preference given to those which control for confounding. We will make this determination based on strength of evidence ratings of insufficient, where there is typically only one study, possibly two small studies for a prioritized (primary) outcome). In this case we will conduct separate searches to identify cohort studies for that specific question and outcome. In the case where a systematic review is recent enough to cover the majority of the available evidence for a given question or subquestion, and evaluates a cohesive group of interventions, outcomes and time frames within the scope for this review, we will include the review as the primary evidence. If there are more than two studies published since the review, our preference will be to use the review only to identify eligible studies for this review.

Non-English Language Studies: We will restrict to English-language articles, but will review English language abstracts of non-English language articles to identify studies that would otherwise meet inclusion criteria, in order to assess for the likelihood of language bias.

Literature Search Strategies

Literature Databases: Ovid® MEDLINE®, Embase®, CINAHL®, Cochrane Central Register of Controlled Trials, and Cochrane Database of Systematic Reviews will be searched.

Publication Date Range: Searches will be conducted across all key questions, with study dates reaching back to the inception of each database. Searches will be deduplicated and screened for inclusion. Searches will be updated while the draft report is open to public comment, to capture any new publications. Literature identified during the updated search will be assessed by following the same process of dual review as all other studies considered for inclusion in the report. If any pertinent new literature is identified for inclusion in the report, it will be incorporated before the final submission of the report.

Supplemental Evidence and Data for Systematic review (SEADS): Manufacturers and other stakeholders of included drugs and devices will be informed about submitting information relevant to this review using a Federal Register notification. A portal about the opportunity to submit information will be made available on the EHC website.

Hand Searching: Reference lists of included articles will also be reviewed for includable studies.

Process for Selecting Studies

In accordance with the *Methods Guide for Effectiveness and Comparative Effectiveness Review*,[18] we will use the pre-established criteria above to screen citations (titles and abstracts) identified through our searches or SEAD submissions to determine eligibility for full-text review. We will begin by screening randomized controlled trials and noting any potential observational studies for include. Observational studies will be screened if evidence from RCTs alone is insufficient to draw conclusions. To ensure accuracy, any citation deemed not relevant

for full-text review will be reviewed by a second researcher. All citations deemed potentially eligible for inclusion by at least one of the reviewers will be retrieved for full-text screening. Each full-text article will be independently reviewed for eligibility by two team members. Any disagreements will be resolved by consensus. A record of studies excluded at the full-text level with reasons for exclusion will be maintained.

Data Abstraction and Data Management

After studies are selected for inclusion, data will be abstracted into categories that include but are not limited to: study design, year, setting, country, sample size, eligibility criteria, population and clinical characteristics, intervention characteristics, funding, and results relevant to each Key Question as outlined in the previous PICOTS section. Information that will be abstracted that is relevant for assessing applicability will include the description of the source of potential study participants, number of patients randomized relative to the number of patients enrolled, and characteristics of the population, intervention (including process details such as monitoring prior to discharge to the outpatient setting, timing or factors determining re-admission, etc.), and care setting such as outpatient or inpatient, details on the type of outpatient setting (e.g. home, home birthing center) or inpatient setting (e.g. hospital, clinic). All study data will be verified for accuracy and completeness by a second team member.

Assessment of Methodological Risk of Bias of Individual Studies

Methods from the *Methods Guide for Effectiveness and Comparative Effectiveness Review* will be used in concordance with the approach recommended in the chapter, Assessing the Risk of Bias of Individual Studies When Comparing Medical Interventions.[18,19] RCTs will be assessed based on criteria established in the *Cochrane Handbook for Systematic Reviews of Interventions* ,[20] and principles for appraisal as developed by the Cochrane Back and Neck Group.[21] If cohort or case control studies are included, they will be evaluated using appropriate criteria developed by the U.S. Preventive Services Task Force.[17] If systematic reviews are included, we will use the AMSTAR-2 tool to appraise these reviews. Based on the risk of bias assessment, individual included studies will be rated as being "good," "fair," or "poor" quality. Any systematic review with multiple flaws, rated poor quality will not be included as primary evidence.

Studies rated "good" will be considered to have the least risk of bias, and their results will be considered valid. Good-quality studies include clear descriptions of the population, setting, interventions, and comparison groups; a valid method for allocation of patients to treatment; low dropout rates and clear reporting of dropouts; appropriate means for preventing bias; and appropriate measurement of outcomes. Studies rated "fair" will be susceptible to some bias, though not enough to invalidate the results. These studies may not meet all the criteria for a rating of good quality, but no flaw is likely to cause major bias. The study may be missing information, making it difficult to assess limitations and potential problems. The fair-quality category is broad, and studies with this rating will vary in their strengths and weaknesses. The results of some fair-quality studies are likely to be valid, while others may be only possibly valid. Studies rated "poor" will have significant flaws that imply biases of various types that may invalidate the results. They will have a serious or "fatal" flaw in design, analysis, or reporting; large amounts of missing information; discrepancies in reporting; or serious problems in the delivery of the intervention. The results of these studies will be least as likely to reflect flaws in the study design as the true difference between the compared interventions.

Data Synthesis

We will construct evidence tables identifying the study characteristics (as discussed above), results of interest, and quality ratings for all included studies, and summary tables to highlight the main findings. We will review and highlight studies by using a hierarchy-of-evidence approach, where the best evidence is the focus of our synthesis for each key question. Data will be qualitatively summarized in tables, using ranges and descriptive analysis and interpretation of the results.

Meta-analyses, using random effects models, will be conducted to summarize data and obtain more precise estimates where there are at least three studies reporting outcomes that are homogeneous enough to provide a meaningful combined estimate. Study designs will be pooled separately (RCTs vs. observational studies). Data from any included high-quality systematic reviews will be handled individually and not pooled. Meta-analysis results for similar outcomes across study types will be compared and discussed where applicable, see section below for evaluation of bodies of evidence with mixed study designs. The feasibility of a quantitative synthesis will depend on the number and completeness of reported outcomes and a lack of heterogeneity among the reported results. To determine whether meta-analysis could be meaningfully performed, we will consider the quality of the studies and the heterogeneity among studies in design, patient population, interventions, and outcomes, and may conduct sensitivity analyses. The Key Questions are designed to assess the comparative effectiveness and harms by patient demographics, patient characteristics (such as gestational age, parity, uncomplicated pregnancy, prior cesarean delivery, etc.), and CR process details, which we will use subgroup analyses to explore, and will look for pre-planned subgroup analyses conducted within individual studies. We will not exclude studies rated as being poor in quality a priori, but poor-quality studies will be considered to be less reliable than higher quality studies when synthesizing the evidence, particularly if discrepancies between studies are present. Sensitivity analyses will be conducted with and without poor quality studies where possible. Meta-regression may be conducted to explore statistical heterogeneity using additional variables for methodological or other characteristics (e.g., quality, randomization or blinding, outcome definition and ascertainment) given a large enough number of studies (e.g. at least six to ten studies for continuous variables and four studies for categorical variables). Publication bias will be assessed using funnel plots and statistical methods when there are at least 10 studies that can be combined in meta-analysis.

Results will be presented as structured by the Key Questions, and any prioritized outcomes will be presented first.

Grading the Strength of Evidence for Major Comparisons and Outcomes

Outcomes to be assessed for strength of evidence were prioritized based on input from the Technical Expert Panel, these are footnoted and listed in bold in the PICOTS table above. Based on this prioritized list, the strength of evidence for comparison-outcome pairs within each Key Question will be initially assessed by one researcher for each clinical outcome (see PICOTS) by using the approach described in the *Methods Guide for Effectiveness and Comparative Effectiveness Review*.[18] To ensure consistency and validity of the evaluation, the grades will be reviewed by the entire team of investigators for:

- Study limitations (low, medium, or high level of study limitations)

    - Rated as the degree to which studies for a given outcome are likely to reduce bias with study design and study conduct, based on risk of bias assessments.
- Consistency (consistent, inconsistent, or unknown/not applicable)
    - Rated by degree to which studies find similar magnitude of effect (i.e., range sizes are similar) or same direction (i.e., effect sizes have the same sign)
- Directness (direct or indirect)
    - Rated by degree to which evidence assesses a) comparison of interest, b) in the population of interest, and measures the specific outcome of interest.
- Precision (precise or imprecise)
    - Degree of certainty surrounding an effect estimate as it relates to a specific outcome. This may be based on sufficiency of sample size and number of events, and if these are adequate, the interpretation of the confidence interval.
- Publication bias (suspected or undetected)
    - Whether selective publishing of research findings based on favorable direction or magnitude of effects is identified using funnel plots or statistical methods.

The strength of evidence will be assigned an overall grade of high, moderate, low, or insufficient according to a four-level scale by evaluating and weighing the combined results of the above domains:

- High—we are very confident that the estimate of effect lies close to the true effect for this outcome. The body of evidence has few or no deficiencies. We believe that the findings are stable, i.e., another study would not change the conclusions.
- Moderate—we are moderately confident that the estimate of effect lies close to the true effect for this outcome. The body of evidence has some deficiencies. We believe that the findings are likely to be stable, but some doubt remains.
- Low—we have limited confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has major or numerous deficiencies (or both). We believe that additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.
- Insufficient—we have no evidence, we are unable to estimate an effect, or we have no confidence in the estimate of effect for this outcome. No evidence is available or the body of evidence has unacceptable deficiencies, precluding reaching a conclusion.

For observational study evidence, the strength starts at moderate for harms outcomes, and low for benefit outcomes. In accordance with the AHRQ Methods Guidance, if the studies have no limitations (i.e. are low risk of bias) this evidence can be upgraded based on one of three conditions. These are: a dose-response is seen, a large magnitude of effect is seen (odds ratio >3), or the effect goes in the opposite direction of plausible confounders, after these have been considered/controlled. In cases where both RCTs and observational studies are included for a given intervention-outcome pair, we follow the additional guidance on how to weight RCTs over observational studies, how to assess consistency across the two bodies of evidence, and how to come to a final rating.[22]

Assessing Applicability

Applicability will be assessed in accordance with the AHRQ's Methods Guide,[18,23] which is based on the PICOTS framework. Applicability addresses the extent to which outcomes associated with an intervention are likely to be similar across the individual studies, bodies of evidence, and individual patients in clinical practice based on different populations, interventions, comparisons, and outcome measures in various settings.[24] For example, lack of inclusion of low-income patients, or those with low access to healthcare, reduces applicability to many clinical practices where an outpatient CR may be considered. Inclusion of only very low-risk pregnancies also reduces applicability to women with moderate risk, who may be candidates for CR depending on the specific risk, method of CR, and monitoring available.

Factors that may affect applicability which we have identified a priori include narrowly defined eligibility criteria and resulting characteristics of included patients, such as demographics (including maternal age, gestational age, race and ethnicity), pregnancy risk factors (such as diabetes, high blood pressure, pre-eclampsia), obstetric factors (e.g. parity), maternal pre-pregnancy health status, including mental health, and intangibles such as birth plan/philosophy and type of provider. Intervention-related factors that may limit applicability include dose and re-administration schedule variation with medications, and balloon-fill volume variation with catheters. In this review, the setting is the key comparison – inpatient versus outpatient – but other features of setting are expected to affect applicability of the findings. These include provider type (e.g., midwife, nurse, or generalist Obstetrician), rural vs. metropolitan, planned home-birth vs. planned inpatient birth, and country. We will use this information to assess the situations in which the evidence is most relevant and to evaluate applicability to real-world clinical practice in typical U.S. settings, summarizing applicability assessments qualitatively, according to the PICOTS framework.

Contextual Question

We plan to follow the methods of the US Preventive Services Task Force to evaluate the contextual question.[17] A targeted search will be designed by a medical librarian with experience in searching for contextual question evidence for USPSTF reviews, including searching for systematic and narrative reviews. The team will also identify any information relevant to this question opportunistically, while reviewing comprehensive literature searches for key questions. The information on the contextual questions will be summarized in the introduction of the report, and discussed in relation to the systematic review evidence on the Key Questions in the Discussion sections.

V. References

1. American College of Obstetrics and Gynecology. Practice Advisory: Clinical guidance for integration of the findings of The ARRIVE Trial: Labor Induction versus Expectant Management in Low-Risk Nulliparous Women. 2018. https://www.acog.org/Clinical-Guidance-and-Publications/Practice-Advisories/Practice-Advisory-Clinical-guidance-for-integration-of-the-findings-of-The-ARRIVE-Trial?IsMobileSet=false. Accessed December 18, 2019.
2. Committee on Obstetric Practice, Society for Maternal-Fetal Medicine. ACOG Committee Opinion No. 764: Medically Indicated Late-Preterm and Early-Term Deliveries. Obstet Gynecol. 2019 Feb;133(2):e151-e5. doi: 10.1097/AOG.0000000000003083. PMID: 30681545.
3. American College of Obstetricians Gynecologists. Practice bulletin no. 146: Management of late-term and postterm pregnancies. Obstet Gynecol. 2014 Aug;124(2 Pt 1):390-6. doi: 10.1097/01.AOG.0000452744.06088.48. PMID: 25050770.
4. World Health Organization. WHO recommendations for induction of labour. Geneva: World Health Organization; 2011. https://apps.who.int/iris/handle/10665/44531 . Accessed December 18, 2019.
5. Grobman WA, Rice MM, Reddy UM, et al. Labor Induction versus Expectant Management in Low-Risk Nulliparous Women. N Engl J Med. 2018 Aug 9;379(6):513-23. doi: 10.1056/NEJMoa1800566. PMID: 30089070.
6. Martin JA, Hamilton BE, Osterman MJK, et al. Births: Final Data for 2017. National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System. 2018 Nov;67(8):1-50. PMID: 30707672.
7. Oregon Health Care Quality Corporation. Oregon Maternal Data Center. http://www.q-corp.org/reports/omdc. Accessed December 18, 2019.
8. Pierce S, Bakker R, Myers DA, et al. Clinical Insights for Cervical Ripening and Labor Induction Using Prostaglandins. AJP Reports. 2018;8(4):e307-e14. doi: 10.1055/s-0038-1675351.
9. Ananth CV, Wilcox AJ, Gyamfi-Bannerman C. Obstetrical interventions for term first deliveries in the US. Paediatr Perinat Epidemiol. 2013 Sep;27(5):442-51. doi: 10.1111/ppe.12068. PMID: 23930780.
10. Lydon-Rochelle MT, Cardenas V, Nelson JC, et al. Induction of labor in the absence of standard medical indications: incidence and correlates. Med Care. 2007 Jun;45(6):505-12. PMID: 17515777.
11. Bernardes TP, Broekhuijsen K, Koopmans CM, et al. Caesarean section rates and adverse neonatal outcomes after induction of labour versus expectant management in women with an unripe cervix: a secondary analysis of the HYPITAT and DIGITAT trials. BJOG. 2016 Aug;123(9):1501-8. doi: 10.1111/1471-0528.14028. PMID: 27173131.
12. Bartha JL, Romero-Carmona R, Martinez-Del-Fresno P, et al. Bishop score and transvaginal ultrasound for preinduction cervical assessment: a randomized clinical trial. Ultrasound Obstet Gynecol. 2005 Feb;25(2):155-9. PMID: 15660437.
13. Carlson NS, Neal JL, Tilden EL, et al. Influence of midwifery presence in United States centers on labor care and outcomes of low-risk parous women: A Consortium on Safe Labor study. Birth. 2019;46(3):487-99. doi: 10.1111/birt.12405. PMID: 30414200.
14. Neal JL, Carlson NS, Phillippi JC, et al. Midwifery presence in United States medical centers and labor care and birth outcomes among low-risk nulliparous women: A Consortium on Safe Labor study. Birth. 2018 Nov 11doi: 10.1111/birt.12407. PMID: 30417436.
15. ACOG Committee on Practice Bulletins -- Obstetrics. ACOG Practice Bulletin No. 107: Induction of labor. Obstet Gynecol. 2009 Aug;114(2 Pt 1):386-97. doi: 10.1097/AOG.0b013e3181b48ef5. PMID: 19623003.
16. Vogel JP, Osoti AO, Kelly AJ, et al. Pharmacological and mechanical interventions for labour induction in outpatient settings. Cochrane Database Syst Rev. 2017(9) PMID: 00075320-100000000-06278.

17. US Preventive Services Task Force Procedure Manual. Rockville, MD: Agency for Healthcare Research and Quality; October 2018. https://www.uspreventiveservicestaskforce.org/Page/Name/procedure-manual. Accessed December 18, 2019.
18. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD: Agency for Healthcare Research Quality; Jan 2014. https://effectivehealthcare.ahrq.gov/topics/cer-methods-guide/overview Accessed December 18, 2019. PMID: 21433403.
19. Viswanathan M, Patnode CD, Berkman ND, et al. Assessing the risk of bias in systematic reviews of health care interventions. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville, MD: Agency for Healthcare Research and Quality; 2017. PMID: 30125066.
20. Higgins J, Savović J, Page M, et al. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., eds. Cochrane Handbook for Systematic Reviews of Interventions version 6 (updated July 2019). Cochrane; 2019. www.training.cochrane.org/handbook.
21. Furlan AD, Malmivaara A, Chou R, et al. 2015 updated method guideline for systematic reviews in the Cochrane Back and Neck Group. Spine (Phila Pa 1976). 2015;40(21):1660-73. PMID: 26208232.
22. Hannan EL. Randomized Clinical Trials and Observational Studies: Guidelines for Assessing Respective Strengths and Limitations. JACC Cardiovasc Interv. 2008 Jun;1(3):211-7. doi: 10.1016/j.jcin.2008.01.008. PMID: 19463302.
23. Atkins D, Chang SM, Gartlehner G, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011 2011/11/01/;64(11):1198-207. doi: 10.1016/j.jclinepi.2010.11.021. PMID: 21463926.
24. Murad MH, Katabi A, Benkhadra R, et al. External validity, generalisability, applicability and directness: a brief primer. BMJ Evid Based Med. 2018;23(1):17. PMID: 29367319.

VI. Definition of Terms

Table 2. Abbreviations

Abbreviation	Definition
ACOG	American College of Obstetricians and Gynecologists
AF	Analytic framework
CD	Cesarean delivery
CDC	Centers for Disease Control and Prevention
CR	Cervical ripening
EPC	Evidence-based Practice Center
FDA	Federal Drug Administration
ICU	Intensive care unit
IOL	Induction of labor
KI	Key Informant
KQ	Key Question
L&D	Labor and delivery
OHSU	Oregon Health & Science University
PICOTS	Population, Intervention, Comparator, Outcome, Setting, Study Design
RCT	Randomized controlled trial
ROM	Rupture of membrane
TEP	Technical Expert Panel
TOO	Task Order Officer
USPSTF	U.S. Preventive Services Task Force

VII. Summary of Protocol Amendments

If needed, protocol amendments will be added in the future.

VIII. Review of Key Questions

The Agency for Healthcare Research and Quality (AHRQ) posted the key questions on the AHRQ Effective Health Care Website for public comment. The Evidence-based Practice Center (EPC) refined and finalized the key questions after review of the public comments, and input from Key Informants and the Technical Expert Panel (TEP). This input is intended to ensure that the key questions are specific and relevant.

IX. Key Informants

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

Key Informants must disclose any financial conflicts of interest greater than $5,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals are invited to serve as Key Informants and those who present with potential conflicts may be retained. The AHRQ Task Order Officer (TOO) and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

X. Technical Experts

Technical Experts constitute a multi-disciplinary group of clinical, content, and methodological experts who provide input in defining populations, interventions, comparisons, or outcomes and identify particular studies or databases to search. The Technical Expert Panel is selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicting opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, study questions, design, and methodological approaches do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and suggest approaches to specific issues as requested by the EPC. Technical Experts do not do analysis of any kind; neither do they contribute to the writing of the report. They do not review the report, except as given the opportunity to do so through the peer or public review mechanism.

Members of the TEP must disclose any financial conflicts of interest greater than $5,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or

content expertise, individuals are invited to serve as Technical Experts and those who present with potential conflicts may be retained. The AHRQ TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

XI. Peer Reviewers

Peer reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodological expertise. The EPC considers all peer review comments on the draft report in preparing final report. Peer reviewers do not participate in writing or editing of the final report or other products. The final report does not necessarily represent the views of individual reviewers. The EPC will complete a disposition of all peer review comments. The disposition of comments for systematic reviews and technical briefs will be published 3 months after publication of the evidence report.

Potential Peer Reviewers must disclose any financial conflicts of interest greater than $5,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers with any financial conflict of interest greater than $5,000 will be disqualified from peer review. Peer reviewers who disclose potential business or professional conflicts of interest can submit comments on draft reports through the public comment mechanism.

XII. EPC Team Disclosures

EPC core team members must disclose any financial conflicts of interest greater than $1,000 and any other relevant business or professional conflicts of interest. Direct financial conflicts of interest that cumulatively total more than $1,000 will usually disqualify EPC core team investigators.

XIII. Role of the Funder

This project was funded under Contract No. HHSA2902015000091 from the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services, through funds provided by a partnership with the Patient-Centered Outcomes Research Institute (PCORI). The AHRQ Task Order Officer reviewed the EPC response to contract deliverables for adherence to contract requirements and quality. The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by either the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.

XIV. Registration

This protocol will be registered in the international prospective register of systematic reviews (PROSPERO).