Chapter 2. Methodology

Technical Advisory Panel

A technical advisory panel (Appendix A) was assembled to provide input from patients, clinicians, and payers to ensure that the scope of the project addressed clinical questions and issues that arise in everyday practice. The panel included obstetricians, family physicians, nurse midwives, payers, and patients. This panel and our national experts and partners provided ongoing assistance throughout the project.

Analytic Framework and Key Questions

Analytic Framework

The analytic framework (Figure 1) represents the strategy we used to organize topic areas and guide the literature search. We developed this framework after a preliminary review of the literature, discussion with local experts, and discussion with national experts.

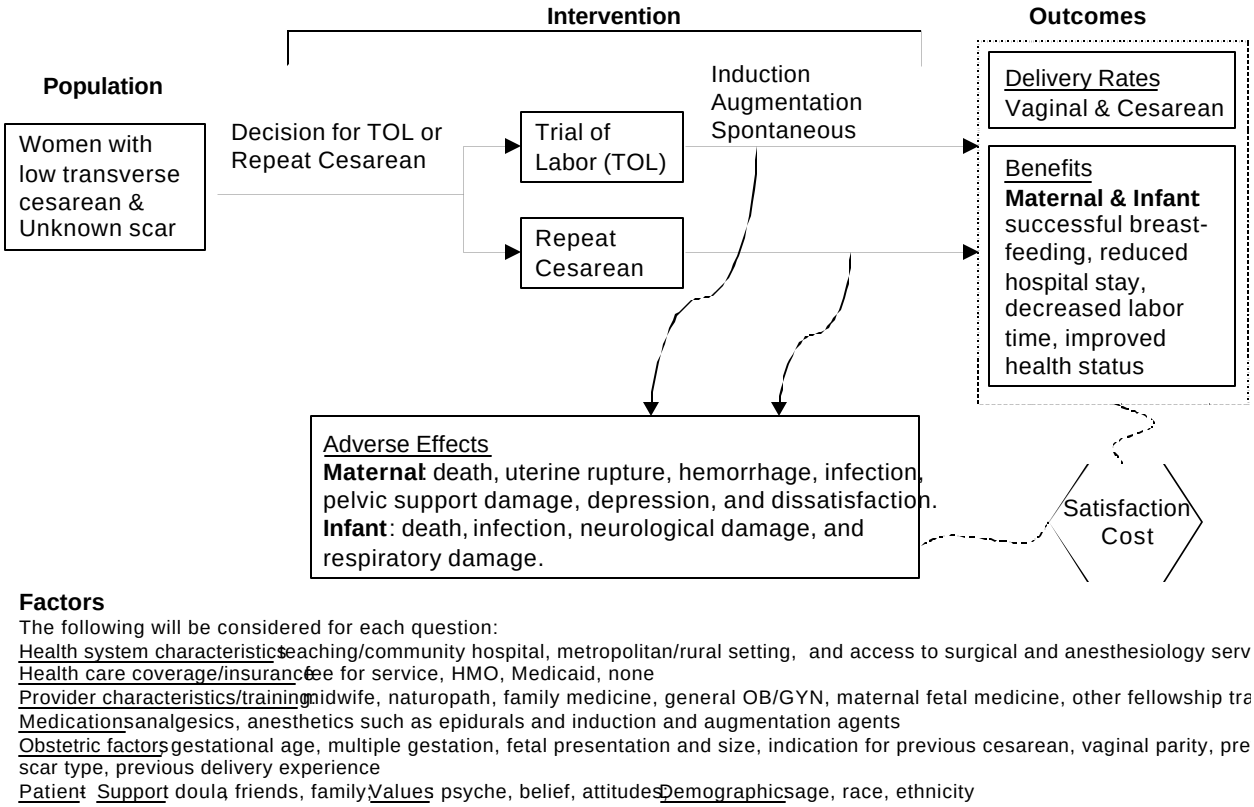
The patients of interest in this report are women with a low transverse cesarean or unknown scar (Figure 1). A woman deciding between having a trial of labor and a cesarean delivery may weigh the benefits and risks, for the mother and the infant, of each approach. A patient who attaches some intrinsic value on the experience of a vaginal birth will be interested in knowing the rate of vaginal delivery. Figure 1 also lists other outcomes and risks ("Adverse Effects") that may be affected by the route of delivery.

All of the benefits and risks listed in the figure may be affected by the method of delivery. However, only some of the risks, such as uterine rupture and, possibly, infant death and damage, are thought to be influenced by having had a prior cesarean section. In defining the scope for this review, we emphasized the benefits and risks that have been reported in studies that included women who have had a previous cesarean delivery. Comparisons of outcomes purely between vaginal and cesarean delivery, but not specifically about VBAC or repeat cesarean delivery, such as breastfeeding, incontinence[12, 13] pelvic support disorders, or infant respiratory sequelae[14] were not considered. Though these are outside the scope of this report, they are certainly important to a woman in deciding between attempted vaginal or cesarean delivery.

The strength and suitability of the evidence regarding the risks of major maternal and infant morbidity and mortality associated with VBAC is the main focus of this report. In judging the suitability of evidence, we took the perspective that the first thing a decisionmaker would want to know is whether the risk of these complications is higher for a trial of labor versus an elective cesarean delivery, *under optimal conditions of care.* That is, the most relevant evidence would compare the outcomes and risks of a properly managed trial of labor to that of a properly conducted elective cesarean delivery. From this perspective, a study comparing the results of VBAC and ERCD that provided little or no information about the quality or content of obstetric care, or that occurred so long ago that the quality of care would be considered poor by today's standards, has little value for patients who are cared for by clinicians who are capable of providing high-quality, up-to-date care.

Some components of obstretric care, as well as some aspects of the setting of this care, might increase the risks of TOL or ERCD. For example, it has been hypothesized that the use (or misuse) or drugs for induction and augmentation might increase the risk of uterine rupture in patients who have had a prior cesarean delivery.[4, 15] Various factors that might affect the outcomes and adverse effects of a trial of labor or an ERCD are listed in Figure 1. We examined the strength of evidence that these factors influence these outcomes and adverse effects and to what extent these factors can explain the results of observational studies of VBAC complications.

Figure 1. Vaginal Birth After Cesarean (VBAC) - Analytic Framework



Key Questions

We addressed two types of key questions. The first group (Questions 1- 7) compares the outcomes of a TOL and an ERCD:

Question 1. What is the frequency of vaginal delivery in women who undergo a TOL (spontaneous onset, induced, and augmented) after prior low transverse cesarean or unknown scar?

Question 2. How accurate are risk assessment tools for identifying patients who will have a vaginal delivery after a TOL?

Question 3. What are the relative harms associated with a TOL (spontaneous onset, induced and augmented) and repeat cesarean?

Question 4. What is the incidence of uterine rupture, and are there methods for preventing major maternal and infant morbidity or mortality due to uterine rupture?
Question 5. What are the health status and health-related quality of life for VBAC and repeat cesarean patients?
Question 6. Regarding VBAC and repeat cesarean, what factors influence patient satisfaction/dissatisfaction with their childbirth experience?
Question 7. How are economic outcomes related to VBAC, repeat CD, and their respective complications?

The second group (Questions 8-10) concern factors influencing the decision to have a TOL:

Question 8. What individual factors influence route of delivery?
Question 9. What factors influence a patient's decisionmaking regarding VBAC or ERCD?
Question 10. How do legislation, policy, guidelines, provider characteristics, insurance type, and access to care affect health outcomes for VBAC candidates?

Literature Search and Selection of Articles

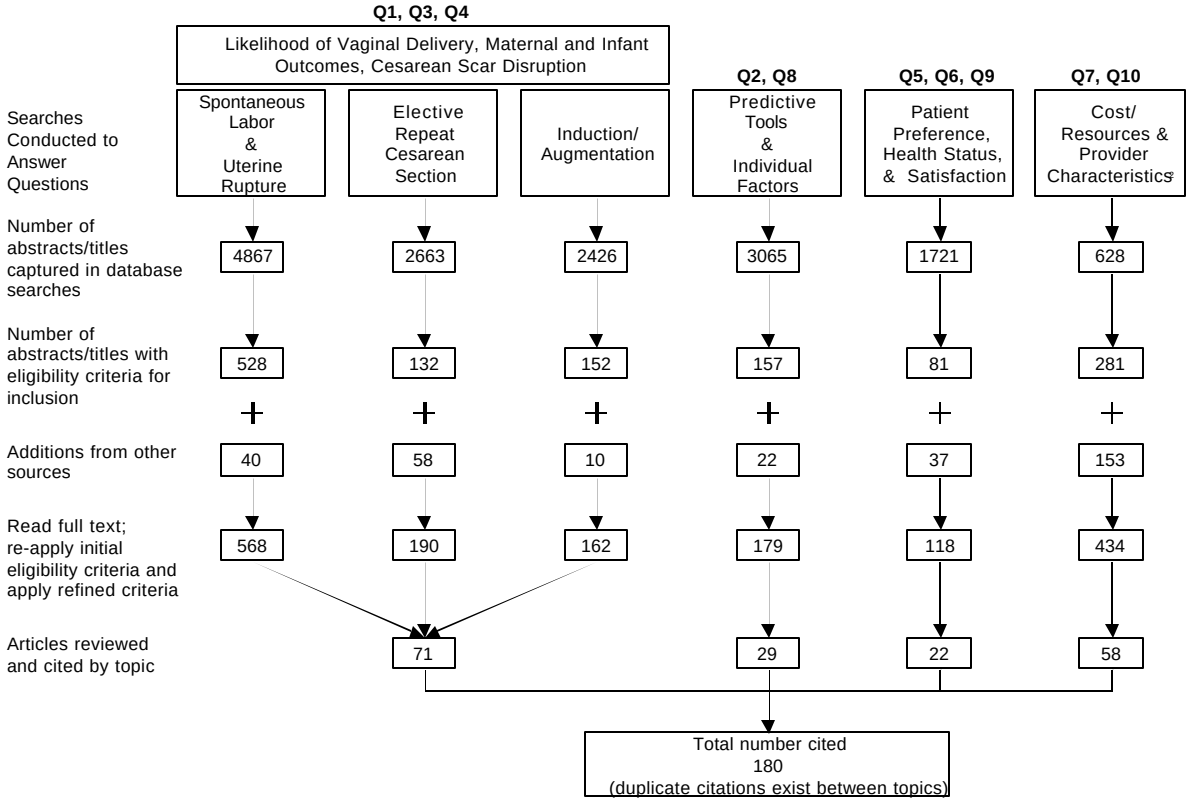
Relevant studies were identified from multiple searches of MEDLINE (1966 to 2002) and HealthSTAR (1975 to 2002), from the reference lists of systematic reviews, and from local and national experts (Appendix A). For relevant literature on specific topics, we also searched the online Cochrane systematic reviews and controlled trials registries, DARE, National Centre for Reviews and Dissemination, and EMBASE databases (Appendix B, search strategies and characteristics).

Databases were searched twice during the course of the project, with the final search in March 2002. Retrieved abstracts were entered into an electronic database (EndNote®). Figure 2 indicates the numbers of abstracts and full-text articles reviewed for all topics in each stage of the review. For all VBAC topics combined, we retrieved 15,370 citations, including 4,867 about spontaneous labor (SL) and uterine rupture; 2,663 about ERCD; 2,426 about induction of labor; 3,065 citations about predictors; 1,721 about patient satisfaction, preference, and health status; and 628 about cost and access.

A lead investigator was assigned for each topic. Two investigators reviewed a random set of titles and abstracts for each topic to select articles for full-text review. When an appropriate level of reliability was reached for inclusion and exclusion of studies, the primary investigator reviewed the rest of the titles and abstracts on the topic. A research assistant tracked the inclusion status and names of reviewers for each abstract reviewed. We retrieved the full text articles of citations that had original data about maternal and infant outcomes relevant to a key question in one or more topic areas.

Studies begun or published before the 1980 National Institute of Health, Consensus Conference on Vaginal Birth after Cesarean, were excluded. The report focused on studies that identified a group of patients with prior cesarean. Studies of the general birthing population were considered if there were no studies that identified patients with prior cesarean. Studies were excluded if they focused on patients with particular conditions such as gestational diabetes, human immunodeficiency virus (HIV), preeclampsia, etc.

Figure 2. Vaginal Birth After Cesarean Section (VBAC): Search and Selection of Citations by Topic



[1]All topics were searched on Medline, Embase , and HealthSTAR . Searches for Induction and Augmentation were also conducted on Cochrane . Searches for Patient Satisfaction, Health Status, and Patient Preference were also conducted on PsychINFO and CINAHL

[2] Includes literature on Economics, Economic Models, Health Services Accessibility, Healthcare providers, Medicaid, Laws, and Guidelines

Exclusions at the title and abstract level were also made for studies that focused on the following: nulliparous patients, vertical, lower vertical, "classical" or "classic" cesarean incision, an inability to differentiate outcomes based upon scar type, vaginal breech delivery, preterm delivery, multiple gestation, or low birth weight. Animal studies, cadaver studies, and studies available exclusively in abstract form were also excluded.

Undeveloped or developing countries were excluded (Appendix C). If the authors described their country as "developing" in either the abstract or the article, it was excluded. Investigators noted this in either the text or evidence tables. Case reports with less than 10 subjects with prior CD were excluded. We also excluded editorials, letters, and nonEnglish language papers.

Case reports, case series, and general population studies (large: n = 100 or greater; small: n = less than 100), were identified but as a rule were not included in the review. Details on suspect or missing data are listed in Appendix D.

When two reviewers disagreed about eligibility, the lead investigator for the topic reexamined the abstract and determined whether the full text of the article should be retrieved. Investigators were encouraged to flag abstracts they believed could be relevant for other topics. Support staff maintained a database to refer these citations to the appropriate investigator if the citations were not already present in the topic-specific abstract database.

After this review, the following were retrieved for full text review: 157 articles about predictors; 528 about TOL and/or uterine rupture; 132 about ERCD; 152 about induction of labor; 81 about patient satisfaction, preference, and health status; and 281 about cost and access.

An additional 320 studies were retrieved after reviewing reference lists of studies and by suggestion of the expert panel or leading researchers in the field. The full texts of these 1,651 studies were retrieved from the library or ordered through inter-library loan. During the abstract review process, 10 VBAC-related systematic reviews were identified and retrieved for review.

Investigators read the full-text version of the retrieved papers and re-applied the initial eligibility criteria. For all topics, we excluded articles if they did not provide sufficient information to determine the methods for selecting subjects and for analyzing data. For some topics, additional criteria were applied to select studies that were systematically reviewed and included in evidence tables as follows.

Included Studies-Evidence Table Level

Data from 180 studies were abstracted and included in the evidence tables described in the results section of this report. Appendix E has details on studies excluded at the paper review level for reasons other than described in the methods section.

Data Extraction

The following information about the patient population, study design, study outcomes, and study quality was extracted from full-text, published studies of VBAC and TOL, induction of labor, ERCD, or uterine rupture, and was used to construct evidence tables: identifying information (study name, years of observation); setting (population-based, referral clinic-based, other); study design (randomized trial, prospective, etc.); interventions (induction, augmentation medications); outcomes studied (infant, maternal, cost, etc.); length of followup; statistical methods for handling confounders (statistical adjustment, stratification, none) and attrition; numbers of subjects recruited, included, and completing study; and characteristics of the sample (demographic variables, number of previous births, other risk factors). For economic evaluations, we also extracted the type of economic evaluation, the primary outcomes reported, data sources, cost unit, discount rate, and what characteristics were varied in the sensitivity analyses and results. Abbreviations and acronyms for study material can be found at the end of the report.

All data were abstracted by the lead investigator for the topic. If the lead investigator encountered difficulty in finding or interpreting information in the published report, a second investigator reviewed the article and a consensus was reached.

Assessment of Study Quality

To assess the internal validity of individual studies, we applied a set of criteria developed by the current United States Preventive Services Task Force and additional criteria developed by the NHS Centre for Reviews and Dissemination, based at the University of York in England. Appendix G shows a detailed description of the quality ratings and tables with quality-rated studies. A brief description of ratings with criteria by study design follows.

RCTs or cohort studies. A study was rated good-quality if it met all the following criteria: comparable groups were assembled initially and maintained throughout the study (followup at least 80 percent); reliable and valid measurement instruments were used and applied equally to the groups; interventions were spelled out clearly; important outcomes were considered;

appropriate attention was given to confounders in analysis; and intention-to-treat analysis was used in RCTs.

A study received a fair rating if any of the following problems were seen: generally comparable groups were assembled initially but some question remained whether some (although not major) differences occurred in followup; measurement instruments were acceptable (although not the best) and generally applied equally; some, but not all, important outcomes were considered; some, but not all, potential confounders were accounted for; and intention-to-treat analysis was used in RCTs.

Studies were given a poor rating if any of the following fatal flaws existed: groups assembled initially were not close to being comparable or were not maintained throughout the study; unreliable or invalid measurement instruments were used or instruments were not applied equally among groups (including not masking outcome assessment); key confounders were given little or no attention; and intention-to-treat analysis was lacking in RCTs.

Case-control studies. A study which met the following criteria was rated good-quality: appropriate ascertainment of cases and nonbiased selection of case and control participants; exclusion criteria applied equally to cases and controls; accurate diagnostic procedures and measurements applied equally to cases and controls; and appropriate attention to confounding variables.

Studies were rated fair if they were recent, relevant, without major apparent selection or diagnostic work-up bias, or accounted for some but not all important confounding variables.

A poor rating was given to a study in this category if it had major selection or diagnostic work-up biases, or inattention to confounding variables.

Economic or cost model studies. For the economic evaluations, Udvarhelyi's[16] ratings were given for six criteria: perspective, benefits, cost data, discounting, sensitivity, and incremental cost-effectiveness ratio (C/E). We assigned to each criterion ratings of good (fulfilled criterion), fair (addressed criterion but not completely or with minor flaw), poor (failed to either address criterion or had a fatal flaw relative to criterion), or not applicable (criterion was not relevant in the context of the evaluation).

Topic Specific Quality Considerations

Investigators were asked to use the study quality ratings as previously described to determine for their topic which quality components were most important in assessing internal validity. This process allowed for some individual topic fit for fatal flaws, etc.

Spontaneous labor and repeat cesarean. To identify which studies to include, we applied a "best evidence" approach.[17] For TOL (SL) and ERCD, we included large population-based and prospective cohort studies. Cohort studies were included because RCTs of delivery method have not been done.

Predictive tools. For this topic, we decided that three of the eight criteria for cohort studies were the most important in determining the quality of each study: (1) comparable groups, (2) clear definition of groups and sufficient description of the distribution of prognostic factors, and (3) consideration of and adjustment for important confounders. Quality was rated as good if all three criteria were met, fair if the groups were comparable and there was adjustment for confounders, and poor if the groups were not comparable or there was no adjustment for confounders.

In addition to the above-mentioned criteria, the evaluation of these diagnostic tests included

several of the factors presented by Reid[18] and Sox,[19] which were: (1) using a prospective study design, (2) avoiding workup or verification bias (i.e., applying the test to all of those eligible for a TOL), and (3) specifying test reproducibility.

Patient satisfaction and health status. Investigators put particular importance on whether the measures for patient health status and psychosocial outcomes were clearly described, including any validation or reliability testing of new health status tools. Specifically, for patient preferences and satisfaction, we put emphasis on methods used to assess patient preferences. Studies that used a method that was independent from the patient's own provider were rated higher than those where the provider assessed this information.

Cost or economic analysis. Specifically for this topic, a poor rating was given for lack of description of the perspective of the economic evaluation, lack of description of the benefits, inclusion of charge data rather than cost data, lack of inclusion of all relevant adverse events, lack of inclusion of discounting (for studies with a time horizon greater than 1 year), lack of sensitivity analyses, and lack of incremental comparisons of alternatives (use of an incremental C/E to compare a more costly alternative to a less costly one).

Access/resources. The studies evaluated were all either databases or cohort studies. The former were typically large national databases and were evaluated using the same criteria as for cohort studies. The main quality criteria used were whether the groups evaluated were comparable at baseline and were controlled for potential confounding variables (including risk adjustment if the groups were not comparable at baseline).

Data Synthesis

Meta-Analytic Methods

Where appropriate, meta-analysis was performed using WinBugs® or StatsDirect® software. To reduce potential bias, only studies of fair or good quality were included in analyses (Appendix G). StatsDirect® was used for comparative studies (e.g., TOL versus ERCD) and WinBugs® was used for noncomparative data (e.g., data for vaginal delivery rates in TOL).

Model estimation using WinBugs® was done using a Bayesian data analytic framework. WinBugs® uses a method of Markov chain Monte Carlo called Gibbs sampling to simulate posterior probability distributions. Noninformative prior probability distributions were used. Absolute risk differences were calculated for each study, and pooled using both random and fixed effects models. Only results from the random effects models are presented, unless these two methods produced significantly divergent results. Statistical heterogeneity was examined. Point estimates using the mean and 95 percent confidence intervals were calculated from 10,000 draws from five Markov chains.

Meta-analysis using StatsDirect® used DerSimonian and Laird random effects methods. The Q statistic tests whether it is reasonable to assume that the treatment effects in the studies to be combined are estimating a single underlying effect size. When the test is significant (e.g., $p < 0.05$) there is significant heterogeneity between the studies' effect sizes. This indicates that the variation seen is greater than that expected from random sampling error. The Q statistic, forest plots and any statistical pooling were done using the StatsDirect® software package (CamCode, England). Where statistically significant heterogeneity was found, pooling was not undertaken.

Individual Factors

Data extraction and data entry were performed using Microsoft Excel 2000®. Because of the nature of this topic and the need for confounding consideration, further analysis involving the calculation of summary estimates using random effects modeling was not considered. Adjusted odds ratios (ORs) for the likelihood of VBAC from each study formed the basis for evaluation. In the situation where the study provided adjusted OR for the likelihood of a failed TOL, the inverse ratio was taken, to approximate the OR for the likelihood of VBAC.