Chapter 4. Conclusions

This report found that there were no high quality data providing definitive answers for decisionmaking about future childbirth following cesarean delivery, one of the most commonly performed surgical procedures in the U.S. (affecting up to 640,000 women each year).

The following summarizes the type of study design, the quality of the evidence from studies, and the suitability of the study design to answer the particular question for each key question.

Summary of Evidence for Key Questions

Key Question	Study Type*	Quality of Evidence	Suitability of Study Design†
Question 1 *What is the frequency of VD in those who undergo a TOL (SL, I, and A) after prior LTC or unknown scar?*	II-2	Fair-Good: Several large prospective and retrospective studies; mostly consistent findings.	Greatest
Question 2 *How do risk assessment tools identify who will have a VD after a TOL?*			
Predictive tools	II-2	Fair-Good: Large cohort studies suggest tools can provide additional data predicting likelihood of (VD).	Greatest
Imaging modalities	I	Good: RCT demonstrated that imaging was ineffective to predict VD.	Greatest
Question 3 *What are relative harms associated with TOL (SL,I and A) and repeat cesarean?*	II-2	Fair-Poor: Many large cohort studies inconsistently defined outcomes.	Moderate
Maternal Death		Fair: Studies consistently found no maternal death risk increase from TOL versus ERCD.	Least
Hysterectomy		Fair-Poor: Many studies failed to report indication for hysterectomy.	Moderate
Transfusion		Fair: Two studies consistently found slightly increased risk for transfusion in TOL although not significant in one.	Moderate
Infection		Poor: Definitions inconsistent.	Moderate
Incontinence/Pelvic Floor		No studies.	Moderate
Infant Death		Poor: Most studies found increased risk of perinatal death for TOL versus ERCD, yet magnitude varied greatly.	Least
Neurologic impairment		Poor: Few studies of poor quality.	Least
Respiratory impairment		No studies.	Moderate

Key Question	Study Type*	Quality of Evidence	Suitability of Study Design†
Question 4 *What is the incidence of uterine rupture of cesarean scar, and are there methods for preventing poor clinical outcomes?*			
Incidence	II-2	Fair-Poor: Several large cohort studies inconsistent in terminology; many with consistent findings of increased risk of symptomatic UR in TOL vs ERCD.	Moderate
Methods for preventing poor outcomes	II-3	Poor: Few studies, variation in case definition. Fetal bradycardia frequently associated with UR; inclusion of fetal tracing findings in definition of UR makes assessing true value difficult.	Least
Question 5 *What are the health status and health-related quality of life for VBAC and repeat cesarean patients?*	None	No studies of women with prior CD.	NA
Question 6 *Regarding VBAC and ERCD,what influences patient satisfaction/ dissatisfaction with the birth experience?*	III	Fair: Two cross-sectional studies with varied findings.	Least
Question 7 *How are economic outcomes related to VBAC, repeat CD, and their respective complications?*	Econ	Fair-Good: One good economic model suggests VBAC cost-effective, provides higher quality of life when chance of VD is 76 percent or greater.	Greatest
Question 8 *What individual factors influence route of delivery?*	II-2	Fair-Poor: Several retrospective cohort studies conducted; all vary in items considered, each with limited adjustment for confounders.	Moderate
Question 9 *What factors influence a patient's decision making regarding VBAC or ERCD?*	I, II, III	Fair: One good RCT and eight fair quality cohort or cross-sectional studies found women who preferred TOL more likely to be White, value process of labor, value social motives such as ease of recovery.	Moderate
Question 10 *How do legislation, policy, guidelines, provider characteristics, insurance type, and access to care affect health outcomes for VBAC candidates?*			
Legislation	II-3	Poor: Few studies only examined impact on VBAC rates, not safety. None examined malpractice rate crisis' impact on access or safety.	Moderate

Key Question	Study Type*	Quality of Evidence	Suitability of Study Design†
Question 10 (continued)			
Guidelines	I, II	Fair-Good: Several studies consistently found the provision of guidelines especially with recommendations of opinion leaders increased VBAC rates; no studies on safety.	Moderate
Provider Characteristics	II	Poor: Several studies, none of which adjusted for differences in baseline risk or potential confounders.	Moderate
Hospital	II	Fair: Consistently found teaching hospitals had higher VBAC rates; no comparisons for safety.	Moderate
Insurance	II	Fair: Several studies with conflicting findings.	Moderate

*Study design categories—I: randomized, controlled trials; II-1: controlled trials without randomization; II-2: cohort or case-control; II-3: multiple time series; III: opinions, descriptive epidemiology. U.S. Preventive Services Task Force (1996).

†Suitability of study design categories—Greatest: For comparison studies: Concurrent comparison groups and prospective measurement of exposure and outcome; For rates: population-based or multicenter prospective cohort studies. Moderate: All retrospective designs or multiple pre or post measurements but no concurrent comparison group; Least: Single pre and post measurements, no concurrent comparison group or exposure, outcome measured in a single group at the same point in time. Community Preventive Services Task Force (2000).

Likelihood of Vaginal Delivery

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

Rates of vaginal delivery for women attempting TOL ranged from 60 to 82 percent. The largest population-based study reported a rate of 60.4 percent. These data may be the best reflection for vaginal delivery rates for the general population who attempt a TOL with low transverse scar across a diversity of settings of care and practice management. The combined vaginal delivery rate for all prospective cohort studies, largely conducted in university and tertiary care settings, was 75.9 percent. Further studies that investigate the true prevalence of vaginal delivery, accounting for practice variation, are needed.

There was a 10 percent reduction in the likelihood of vaginal delivery when oxytocin was used for ether induction or augmentation. There was a similar trend in reduced likelihood of vaginal delivery with prostaglandins. Most studies did not report rates for patients requiring medical augmentation or induction of labor separately from patients undergoing spontaneous labor. Furthermore, studies that did report separate rates, were not able to account for the contribution of reason for augmentation or induction, nor the impact of practice variation. Leaving insufficient data to determine the effect of medical induction and augmentation of labor.

Predictive Tools

How accurate are risk assessment tools for identifying patients who will have a vaginal delivery after trial of labor?

In considering whether to attempt a TOL or ERCD, patients, clinicians, payors, and policy-makers are confronted with the dilemma of weighing the likelihood of probabilities for vaginal delivery and health outcomes for each option.

Two validated scoring systems were identified.[36, 40] These two scoring systems shared the design of incorporating various predictive factors available at a patient's admission, similar study patient exclusion criteria (e.g., classical or low vertical incision, multiple gestations, and malpresentation), and a roughly similar range of predicted vaginal delivery probabilities of 45 to 95 percent. In addition to these similarities, the two scoring systems also shared several limitations. First, both scoring systems were based on preselected populations of patients who were willing to attempt a TOL. Because of this design, both studies are affected by verification or workup bias, where the results are relatively distorted by the fact that not everyone who is eligible for a TOL is included in the study (e.g., the patient who is eligible for a TOL, but decides to have a ERCD is not incorporated into the study and not used for the creation of the scoring system). Another common limitation is that these scoring systems were created and validated for use at the time of admission, thus invalidating the application of the scoring systems at any other point during the pregnancy. For example, Flamm stated that because cervical dilation and effacement often change dramatically between the last prenatal examination and the time of admission, the use of his scoring system before the onset of labor would yield an incorrect prediction. The last common limitation stems from the included predicting variables themselves such as accuracy of a patient's past obstetric history (e.g., indication of a prior CD) if the medical record is not available, and the variable and subjective in nature of cervical dilation and effacement. The lack of accurate past obstetric data or the variability of various clinical findings between providers could potentially affect the precision of the predicted results.

However, beyond these similarities lie several differences that make the Flamm scoring system a relatively better predictive tool. First of all, Flamm's scoring system was developed prospectively and with a considerably larger sample size, compared with the Troyer scoring system (2,502 and 264, respectively). Flamm's scoring system can also be said to be more precise and accurate, in that the point values assigned to each of the included variables were based on the Beta coefficients of the logistic regression model. This system, which was not employed by Troyer, takes into account the relative predictive weights for each variable, while controlling for any possible confounding distortion. The use of a 10-point scoring system by Flamm also increases the accuracy and precision of his system by allowing for a more exact prediction of the probability of success, relative to Troyer's four-point scoring system. The value of a scoring system depends on its ability to accurately stratify patients into high and low-risk groups with low false positive or negative rates. In the case of TOL, an ideal tool would stratify all women eligible for a trial of labor into those with high and low likelihoods of vaginal delivery, with minimal false positives. The tool should minimize the number of patients predicted to be at high chance for vaginal delivery that actually have to have a cesarean after a lengthy trial of labor (false positives), because it is this group that has the highest risk to sustain complications of TOL such as uterine rupture. Flamm's test was able to provide additional information to slightly under one-half of the population tested, with a relatively low false positive rate of 2.6 percent. In order to know whether this tool is effective, it needs to be tested in different populations with differing baseline VBAC rates, and ideally tested in all eligible women rather than just those who already chose TOL.

Of the seven imaging studies identified, only one received a good quality rating.[47] Although this RCT was similar to the other studies, in that it lacked any statistical adjustment for

confounding, its randomization of subjects presumably allowed for control of confounding through study design. The finding that 60 percent (33/55) of those considered to have an inadequate pelvis by postpartum XRP had a vaginal delivery, compared with the 30 percent (27/89) of those considered to have an adequate pelvis by postpartum XRP, provides support for the conclusion that XRP is a poor predictor of TOL outcome and might unnecessarily increase CD rates.

Maternal and Infant Health Outcomes

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

There is no direct evidence comparing the risks and benefits of TOL relative to ERCD in similar patients. Several fair and good quality cohort studies provide indirect evidence about the relative benefits and harms associated with each route. Their findings are itemized below:

- Maternal death rates did not differ between TOL and ERCD.
- The best evidence suggests that hysterectomy rates do not differ between TOL and ERCD.[5]
- Rates of infection were increased in ERCD versus TOL (8.6 to 9.73 percent versus 6.6 to 6.79 percent).[5, 24]
- Studies that performed subgroup analyses for TOL with and without vaginal delivery consistently reported that rates of infection were significantly higher in women who had a TOL but ultimately had a cesarean delivery.
- There is conflicting evidence regarding whether induction of labor had any effect on infection rates.
- There is insufficient evidence regarding the effect of TOL and ERCD on APGAR score and respiratory morbidity.
- No study measured infant death directly attributable to a mother's choice of TOL or repeat CD.
- Two large population-based studies report increased risk of perinatal death associated with TOL, but they differ in the magnitude of risk.(90/10,000 TOL versus 50/10,000 ERCD[5] compared with 12.9/10,000 TOL versus 1.1/10,000 ERCD.[6])

Methodologic deficiencies in the literature are striking. Comparisons across studies were hampered by lack of standards for reporting severity of disease or condition, and inconsistencies in definitions of outcomes. Studies often did not pay close attention to comparability of groups, specifically, the ERCD group was often not ensured to be otherwise eligible for TOL. Other factors such as parity, type and number of previous cesarean, were often not considered.

Studies did not pay close attention to and account for the importance of co-interventions such as use of oxytocin and other medical agents for augmentation or induction of labor.

Most importantly, variations in reporting of important clinical outcomes such as hysterectomy, infection, maternal mortality, and perinatal mortality made it difficult to determine true probability of outcomes, potential preventive measures, or outcomes that were directly attributable to route of delivery or labor management. Lack of precision made it difficult to

determine whether the rates truly represented risk of clinically significant outcomes or significant misclassification or confounding.

There were no studies of the long-term consequences of TOL versus ERCD such as incontinence, pelvic support disorders, or infant sequelae from neurologic or respiratory disorders.

Uterine Rupture

What is the incidence of uterine rupture, and are there methods for preventing major morbidity and mortality due to uterine rupture?

Studies varied in their use of terms to describe the spectrum (e.g., asymptomatic, symptomatic, clinically significant) of uterine rupture of the cesarean scar. Our best attempt to separate the groups in a meaningful way found that there was no difference in rates of asymptomatic uterine rupture (dehiscence) between TOL and ERCD. There was a significant increase in the occurrence of symptomatic uterine ruptures in TOL. Specifically, for every 10,000 women attempting TOL there would be 27 additional symptomatic uterine ruptures. Based on the frequency and severity of symptomatic rupture, for every 10,000 women undergoing a trial of labor, there would be 1.5 uterine rupture related perinatal deaths and 4.8 rupture related hysterectomies.

Lack of precise definitions also prevents the ability to determine the value of certain premonitory signs. Because the definition of uterine rupture frequently includes ruptures discovered when cesarean is performed for fetal heart tracing disturbances, it is not possible to determine the accuracy of fetal tracing as a premonitory sign.

Health Status

What are the health status and health-related quality of life for VBAC and repeat cesarean patients?

No studies provide information on health status or health-related quality of life, related to TOL versus ERCD.

Patient Satisfaction

Regarding VBAC and repeat cesarean, what factors influence patient satisfaction/ dissatisfaction with their childbirth experience?

It is important not only to consider the health outcomes for TOL and VBAC, but also whether patients are satisfied with their childbirth experience. Only two fair cross-sectional studies provided results on satisfaction for women attempting VBAC or ERCD. Other studies allowed the patient's provider to measure satisfaction, introducing the possibility of measurement bias.

Cost and Health Care Resources

Cost

Discussion of economic evaluations. The use of cost per QALY from a societal perspective as an economic outcome to compare health care delivery options is recommended by current guidelines.[100] While there is no single threshold value for cost per QALY in the US, the upper limit of cost effectiveness of $50,000 per QALY used by Chung et al. is a reasonable limit for the US health care system.[87] This limit can reflect one extra QALY at a cost of $50,000 or 50 extra QALYs at a cost of $1,000 per QALY. A value of $50,000 per QALY is slightly less than the cost per QALY for treatment guided by routine coronary angiography compared with initial medical therapy without angiography, or use of driver-side and passenger-side airbags compared with driver-side air bags alone.[174]

The use of QALYs as an economic outcome for methods of delivery means that both the mother and the newborn contribute QALYs to the analysis. It seems appropriate that both maternal and newborn QALYs should be counted, as both are outcomes influenced by the decision on mode of delivery. Economists typically do not differentiate QALYs on the basis of the age of the person receiving the QALY. That is, a QALY is counted the same for a senior age 80 as for a child age 5. Thus, a comparison between a childhood vaccination program and hip replacement surgery is facilitated by using cost per QALY.

Additional analyses using the model of Chung et al.[87] would be useful. The authors could have performed two-way sensitivity analyses with each of the other sensitive variables listed above and TOL success probability to determine how sensitive these results are to two variables at once. For example, if an increase of 0.5 percent in the probability of cesarean rupture were to shift the decision point from 74 to 80 percent, then both of these two factors would need to be predicted to determine which delivery option was more efficient. That is, the results might be sensitive to more than one variable at a time. One problem with the recommendations of this study based on TOL success rate is that the recommendations ignored the imprecision of the estimated TOL success rate. If the TOL probability of success were 72 percent or 76 percent with a prediction error of +/- 4 percent (e.g., a CI for the prediction of 68 percent to 76 percent for a TOL success rate of 72 percent), the prediction interval would include the decision cut point of 74 percent. This means that the prediction does not select an efficient option in this case. A Monte Carlo simulation analysis that would allow introduction of random variation into the model of Chung et al. could help to evaluate the effect of uncertainty in the prediction parameter. For example, instead of using a predicted probability on TOL success, one could use the expected probability and the standard error around the probability to generate a sample of individuals, determine the experience of these individuals, and estimate the resulting cost per QALY. Another concern is the inclusion of fecal and urinary incontinence during the first year after birth in the model of Chung et al. As summarized elsewhere in this report, the evidence for a higher rate of these adverse events in TOL than ERCD is inconclusive. The authors should have included no additional cases of incontinence in the sensitivity analyses.

The valuation of different costs in these economic evaluations needs review. There are a number of costs associated with TOL and ERCD that are very difficult to measure. These events include, but are not limited to, cerebral palsy, loss of fertility after a hysterectomy, or death of the mother or of the newborn. These events have substantial societal costs that might be problematic to measure. To the extent these events are not properly valued in the above analyses,

the analyses are potentially biased. The use of a broad range of sensitivity values might address this concern to some extent. With respect to major neonatal adverse events such as cerebral palsy, the costs include more than direct medical costs. The societal costs (e.g., long-term care, special education, lost productivity, and legal costs) of a major neonatal adverse event might be substantially higher than the direct medical costs. For example, the productivity lost for a newborn with a cognitive deficit could be substantial from a societal perspective. However, these societal costs were not included in the model of Grobman et al.[88] Cerebral palsy after uterine rupture had the highest cost in this model (base case about $180,000) but occurred with very low probability. Maternal and neonatal deaths were not explicitly valued except in sensitivity analyses and then with a relatively small value ($100,000), because of the payer or provider perspective. While it is likely that these probabilities change with each subsequent pregnancy (e.g., a successful TOL indicates a higher probability of success for future TOLs).[107] Another problem with costs is the true cost of the perinatal period (including times associated with labor and delivery for a TOL and with surgical processes for RCD). Chung et al. used charges for these costs; charges might not reflect actual time spent in labor and delivery or in surgery. More detailed studies that evaluate these times for series of patients would improve these models. These details are as important as LOS (see next section on health care resources below) for an accurate estimate of total costs.

The model of Chung et al.[87] also considers only one pregnancy. The model of Grobman et al.[88] did include more than one pregnancy after an initial CD. In this latter model, probabilities for each subsequent pregnancy appear to be the same as for the index pregnancy. Some women might be expected to have additional pregnancies and each pregnancy and the modes of delivery in the previous pregnancies are likely to modify the probabilities for subsequent pregnancies. For example, a repeat CD might increase the risk of other adverse events if a TOL is considered for the next pregnancy. Similarly, a successful VBAC means that a woman is more likely to have a TOL end in VBAC for subsequent pregnancies. While the data for subsequent pregnancies might be somewhat limited, the impact on future pregnancies is important.

In summary, the model of Chung et al.[87] provided the best evidence of the relative value of TOL and ERCD, and suggested that the cost-effectiveness of TOL versus ERCD depends strongly on the probability of successful VBAC after a TOL. If this probability is "high," VBAC is more cost-effective, while if this probability is "low," ERCD is more cost-effective. Additional research is needed before precise values of high and low in the above can be assigned. Also there is likely a range of probabilities between the high and low values in which the cost-effectiveness might be indeterminate. The discussion above describes some additional analyses using the model of Chung et al. that might address some of these issues raised. However, other concerns, especially achieving a prediction tool of the desired precision, might be problematic. A second model by Grobman et al.[88] provided only fair evidence, from a payer perspective, of the medical costs of TOL versus ERCD. Thus, Grobman et al. do not provide conclusive evidence of the value of VBAC over ERCD.

Health Care Resources

All studies were rated poor, mainly for lack of adjustment for potential confounding variables.

Individual Factors

What individual factors influence route of delivery?

This review identified 96 studies that met the requirements for inclusion. However, upon further review, 83 of these studies were considered of poor quality and were subsequently removed from the analysis. The most common reason that studies were rated poor was due to lack of adjustment for important confounders. While many studies commented on the extensive list of factors that influence the outcome of TOL, very few studies actually considered those factors when conducting their analyses. Instead of stratifying their analysis or running multivariate models (e.g., logistic regression), studies often provided only bivariate analyses (i.e., Chi-square, Fisher exact, or t-tests). By neglecting to control for confounding, the measures of association provided by these studies might be distortions of the true association and hence should be interpreted with caution.

Overall there was an *increased likelihood of vaginal delivery* for women who had a prior vaginal delivery (particularly VD after cesarean), maternal age less than 40 years, a nonrecurrent indication for one's prior CD, and favorable cervical assessment. There was a *decreased likelihood of vaginal delivery* for women with an increased number of prior CDs, gestational age greater than 40 weeks, birth weight greater than 4000 grams, and augmentation of labor. Although all of these significant findings come from good to fair quality studies, it is important to remember that some of these factors do in fact vary between individual health care providers. For example, the cervical examination performed by one provider may differ from the exam of another or in another instance; the decision to augment a labor and how aggressively this approach should be applied may also be dramatically different between providers. In any case, these inter-provider variations may have not only affected the obtained results and perceived associations, and also has possible implications in the use of such knowledge in the clinical realm.

Patient Preferences

What factors influence a patient's decisionmaking regarding VBAC or ERCD?

A woman's choice for delivery was often based on social motives (e.g., easier recovery so she can care for her baby and children at home). Only four of 11 studies cited safety of the mother or bay as important reasons for delivery choice. It remains unclear if VBAC education increases the proportion of women who choose TOL. Future studies should include education, ideally before next pregnancy.

Provider Characteristics, Legislation, Access to Care

How do legislation, policy, guidelines, provider characteristics, insurance type, and access to care affect health outcomes for VBAC candidates?

One of the things a decisionmaker would want to know in deciding between TOL and ERCD is what conditions of care including practice management, training of the provider, and hospital characteristics increase the risks of each choice. There were no high quality data for this issue, in

fact, studies of these factors exclusively examined VBAC rates rather than the safety of each choice.

Legal or Legislative Factors

No study provided direct evidence for the impact of rising malpractice rates on VBAC or ERCD. Two studies were identified that provided any data regarding legal and/or legislative effects. One study in Florida found a significant difference in VBAC rates before and after enactment of statewide legislation emphasizing dissemination and peer-review enforcement of guidelines. Analysis failed to consider underlying time trend in VBAC rates independent of legislation. Another study in New York found small changes (ORs between 0.95 and 1.0) in probability of VBAC for either hospital-paid loss due to malpractice claims or $5,000 increase in annual physician insurance premium increase. No other studies of the effects of increasing insurance premiums were identified.

Guidelines

- A randomized trial[133] demonstrated that opinion leaders are able to modify provider behavior to a greater extent than audit and peer review.
- A second randomized trial[134] failed to show a significant change in response to audit and peer review.
- Two retrospective cohort studies[135, 136] used data over time to show increases in VBAC rates in response to national VBAC guidelines.

Provider Characteristics

Provider characteristics such as training to perform a cesarean, clinical volume, and management characteristics may affect outcomes of TOL and ERCD. Though these may be important factors, no studies that examined these factors, controlled for important confounders such as patient selection bias. Thus, there is no evidence as which if any of these factors may increase risk.

Hospital Characteristics

- Most studies of the effect of teaching hospitals found that teaching hospitals had higher VBAC rates.
- Studies disagreed whether the presence of a NICU in the hospital affected VBAC rates
- In small rural hospitals, three studies of small case series found VBAC success rates of 67 to 88 percent with no serious adverse events. More extensive experience might modify this result.

Insurance Types

There were conflicting data regarding the impact of types of health insurance on VBAC rates.